

Result No.	C	1	76	Query			DB	ID	Description
				Score	Match	Length			
				843	5.0		28	BZ643413	OGA05W7TM
		2	69.8	4.6			29	BX139987	Danio rer
		3	69.4	4.6			29	FR048073	Fugu rubr
		4	69.4	4.6			29	BX173672	Danio rer

SUMMARIES

ALIGNMENTS

Seq primer: TR

ALIGNMENTS

Seq primer: TR

FEATURES	Class: sheared ends.		Definition	
	Location/Qualifiers		Accession	
	1..843		BX139987	
	/organism="Zea mays"		GI:27971314	
source	/mol_type="genomic DNA"		GSS	
	/strain="B73"		Danio rerio (zebrafish)	
	/db_xref="taxon:4577"		Danio rerio	
	/clone_lib="ZM 0.7 1.5 KB"		Submitted (13-MAR-2003)	
ORIGIN	/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"		The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Unpublished	
	5.0%; Score 76; DB 28; Length 843;		This sequence was generated from the T7 end of BAC 99E7. 99E7 is part of the DanioKey BAC library created by R. Piasterk and N.V. Keygene. Further details:	
	43.7%; Pred. No. 6.6e-07;		http://www.sanger.ac.uk/Projects/D_rerio/.	
	Matches 334; Conservative 0; Mismatches 430; Indels 0; Gaps 0;		Location/Qualifiers	
QY	171 TAATGCCAAATGATCTGATTCGAATAAGATGGAGCTTCATTATTCCTCAAGGTAA 230		1..844	
	778 TACTGCTGCTGATGCTGCTACTGCTGCTGCTACTGCTACTGCTACTGCTACTGCT 719		/organism="Danio rerio"	
	QY 231 AGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATCTCTTTTACGCAACTACTATGG 290		/mol_type="genomic DNA"	
	DB 718 TACTGCTACTACTGCTGCTACTACTACTGCTACTACTGCTACTACTGCTACTGCT 659		/db_xref="taxon:7955"	
QY	291 AGGAGAACATATGTAACATAAAGATGACTGCGCTCAGCAAAATGAGATTGCGACATGGGC 350		/clone="DKEY-99E7"	
	DB 658 TGCTACTACTACTACTGCTACTGCTACTGCTGCTACTGCTGCTACTGCTGCTACTGCT 599		/tissue type="Testis"	
	QY 351 TAAATCTATATACCTAATATACACAGTTTCACATGTTTACTGAGTCAAAATATGATGG 410		/note="vector pindigoBAC-536"	
	DB 598 TGCTACTCCAGCTACTACTGCTGTTGCTGCTACTGCTGCTGCTACTGCTACTGCTGCTAC 539		ORIGIN	
QY	411 TACTGAGGTTTAAATGTTTCCCAATATGGATATTTATTTGTTTCTAGCAGCTGTTATATA 470		Query Match	
	DB 538 TGCTGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479		Best Local Similarity	
	QY 471 TGGAGCTGTATTAATGTTTACATGTAATCCAAATGCTACTATTTCAATGAAAAGATAC 530		46.3%; Pred. No. 1.8e-05;	
	DB 478 TGCTGCTGCAACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTACTAC 419		Matches 230; Conservative 0; Mismatches 267; Indels 0; Gaps 0;	
QY	531 TGATGCGACATGGGGAGATGGTGTGGAAGAACTGTAGATCAAAACGTTACTCGGTTGG 590		137 ATAAAGCATATAAGTTTTTGTATGCGAGAAATAGATATGCAATGATCTGATTGCGAATA 196	
	DB 418 TACTGCTAACTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 359		348 ATAAATTAATTAATGATAATTAATGATAATTAATGATAATTAATGATAATTAATTAATA 407	
	QY 591 TGATACAGTCAAAATATATTAATTAAGATGCAATGCAATGCAATGCAATGCAATGCAATG 650		197 AAGATGGAGCTCTCTTATTTAAATTCCTCAAGGTAAAGAGCTGAGTATAAAGCTTCAACTG 256	
	DB 358 TACTACTGCTACTGCTTCTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAC 299		408 TTGATAATTAATGATAATTAATTAATGATAATTAATGATAATTAATTAATTAATGATA 467	
QY	651 AGTGATCAATATGTTTATAAGGATATATGCCATCTGCTTCTGTAGTTGATTGAAACGA 710		257 ATTTTAAATTCCTTTTACGACAACTACTTAATGCGAGGAGACATATGTAACCTAAAAAG 316	
	DB 298 AGCTACTACAGTACTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 239		468 ATAATAATAATAATGATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATG 527	
	QY 711 AGGTCTTATGAAGTACTATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 770		317 ATACTCGCTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATATCAGCTTAATCTACAC 376	
	DB 238 TACTGCTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTGCT 179		528 ATAATAATAATAATGATAATTAATAATGATAATTAATAATGATAATTAATAATGATAATTA 587	
QY	771 TTCGGAAGAAAGCACTGGGAAGTATAACCTGTTAGAGGAAAATAATAATTTTCAGATTAC 830		377 CAGTTTCCCACTGTTACTGAGTCAAAATTAATGATGCTAGGTTTATTAATGTTTCCCAAT 436	
	DB 178 TGCTACTGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 119		588 ATGATGATAATAATAATGATGATAATAATGATAATAATAATGATAATAATAATAATGATA 647	
	QY 831 TATTTCCTGGGAGCTACCAATCTCCAAACGGGAATATTAACAATCTCAACTCAAGG 770		437 ATGGATATTAATTAATGTTTCTAGCAGCTGTTAATAATGAGCTGTTAATTAATGTTGTTGTTG 496	
	DB 118 TAATACTGCTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTGCT 59		648 ATAATAATAATAATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATG 707	
QY	891 CTTTTTTTATAAGGATAATAACAATCAACAGTCACTTATACAG 934		497 TAACTCCAAATGCTACTATTTTCATGAAAAGAAATACTGATGCGACATGGGAGATGGTGGTG 556	
	DB 58 TACTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15		708 CTAAATGACAATAATAATAATAATGACAATAATAATAATAATAATAATAATAATAATAATG 767	
	QY 891 CTTTTTTTATAAGGATAATAACAATCAACAGTCACTTATACAG 934		557 GAAAAACTGTAGATCAAAAAACGCTACTCGGTTGGTGATACAGTCAAAATATCTATTACTTT 616	
	DB 58 TACTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15		768 ATAATAATGATACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATG 827	
RESULT 2	BX139987		617 ATAGAATGCACTCAAT 633	
	844 bp		828 ATAATAATGATGATAAT 844	
	DNA		494 bp	
	linear		linear	
RESULT 3	BX139987		FR0048073	
	844 bp		GSS 05-JAN-2001	
	DNA		linear	
	linear		linear	

DEFINITION Fugu rubripes GSS sequence, clone 263K15b08, genomic survey sequence.
 ACCESSION AL444858
 VERSION AL444858.1 GI:12052694
 KEYWORDS GSS; genome survey sequence.
 SOURCE Takifugu rubripes (Fugu rubripes)
 ORGANISM Takifugu rubripes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorphia; Tetraodontiformes; Tetraodontidae; Takifugu.
 REFERENCE 1 (bases 1 to 494)
 AUTHORS Clark, M.S.
 TITLE Direct Submission
 JOURNAL Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB, UK Email: biohelp@hgm.mrc.ac.uk
 COMMENT Vector: pBluescript II KS
 PRIMER: KS
 DESCR: One pass dye-terminator sequencing of BAC (pBelobACII) cloned genomic sequence
 The BACs can be obtained from <http://www.incyte.com>.
 FEATURES
 source
 Location/Qualifiers
 1..494
 /organism="Takifugu rubripes"
 /mol_type="genomic DNA"
 /db_xref="taxon:31033"
 /clone="263K15b08"
 /clone_lib="BAC 263K15"
 ORIGIN
 Query Match 4.6%; Score 69.4; DB 29; Length 494;
 Best Local Similarity 48.6%; Pred. No. 2.1e-05;
 Matches 190; Conservative 0; Mismatches 201; Indels 0; Gaps 0;
 QY 236 CTGAGTAAAGCTTCAACTGATTTTAACTTCTCTTTTACGACAACTACTAATGGGGA 295
 DB 22 CTGGCGCTTCTACTACTACTAGTACTGCTGCTACTACTAGCTACTGCTGCTGCT 81
 QY 296 GAACATATGTAAGTAAAGAGTACTGCTGCTAGCAATGAGATGCGACATGGCTAAAT 355
 DB 82 CGACTACTGCGACGACGCTACTACTACTGCTGCTGCTGCTGCTACTACTACTA 141
 QY 356 CTATATCAGTAATPACTACACGATTTCCACTGTTACTAGTCAAAATGATGGTACTG 415
 DB 142 CGAGGCTGCTACTACTACTGCTACTACTACTACTACTACTACTGCTGCTACTA 201
 QY 416 AGTTATTATGTTCCCAATATGGATATTATTATGTTTCTAGCAGCTTTAATAATGGAG 475
 DB 202 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 261
 QY 476 CTGTAATTTATGTTTACATCTGTAATCTCAATGCTACTATTTCATGAAAGAAATGATG 535
 DB 262 CTACTACTACTACTACTACTACTACTGCTACTGCTACTACTACTACTACTACTACTA 321
 QY 536 CGACATGGGAGATGGTGGTGGAAAACTAGATCAAAAAACGCTACTCGGTTGGTGATA 595
 DB 322 CTACTACTACTACTGCTGCTGCTGCTGCTGCTGCTACTACTACTACTACTACTGCTA 381
 QY 596 CAGTCAATATATCTATTACTATTAGATGC 626
 DB 382 CTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 412
 RESULT 4
 BX173672/c
 LOCUS BX173672 829 bp DNA linear GSS 13-MAR-2003
 DEFINITION Danio rerio genomic clone DKEY-150M6, genomic survey sequence.
 ACCESSION BX173672
 VERSION BX173672.1 GI:28005377
 KEYWORDS GSS.

SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 829)
 AUTHORS Humphray, S.J., Huckle, E. and Durham, J.L.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphray@sanger.ac.uk Unpublished
 COMMENT This sequence was generated from the SP6 end of BAC 150M6. 150M6 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
 FEATURES
 source
 Location/Qualifiers
 1..829
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEY-150M6"
 /tissue_type="Testis"
 /note="vector pindigoBAC-536"
 ORIGIN
 Query Match 4.6%; Score 69; DB 29; Length 829;
 Best Local Similarity 45.2%; Pred. No. 2.7e-05;
 Matches 252; Conservative 0; Mismatches 305; Indels 0; Gaps 0;
 QY 385 ACTGTTACTGAGTCAAAATGATGTTGTTACTGAGGTTATTAAATGTTCCCAATATGGATAT 444
 DB 645 AATAATAATAATAAGATAATAATAATAATGATGATGATGATGATGATGATGATGAT 586
 QY 445 TATTATGTTTCTAGCTGTTAATAATGAGCTGTGTAATTAATGTTTACATCTGTAATCCCA 504
 DB 585 AATAATAGTAATAATAATAATAATAATAAGAAATAATAATAATGATGATGATGATGAT 526
 QY 505 AATGCTACTATTTCATGAAAGATGATGATGCGACATGGGAGATGGTGGTGGAAACT 564
 DB 525 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGATGAT 466
 QY 565 GTAGATCAAAAAACGTTACTCGGTTGGTGATACAGTCAAAATATATCTATTACTTATAAGAA 624
 DB 465 GATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 406
 QY 625 GCAGTCAATTTATCATGGTACAGAAAAAGTGTATCAATATGTTTATAAGGATATCTATGCCA 684
 DB 405 GATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGATGAT 346
 QY 685 TCTGCTTCTGTAGTTGATTTGAACGAGGGTCTTTATGAAGTAACTATTACTGATGGATCA 744
 DB 345 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 286
 QY 745 GCGAATATTACAACTCTAACTCAAGGTTGCGAAAAAGCAACTGGGAAGTAACTCTGTTA 804
 DB 285 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 226
 QY 805 GAGGAAAAATAATAATTTTCACGATTACTATTTCGTTGGGCGAGCTACCAATATCTCCCAACCGGA 864
 DB 225 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 166
 QY 865 AATACTCAAAATGGAGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 924
 DB 165 AATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 106
 QY 925 ACTTATACAGGAGTATT 941
 DB 105 TAAAAATAGAAAAATATT 89
 RESULT 5
 CNS02HA4
 LOCUS CNS02HA4 989 bp DNA linear GSS 01-SEP-2000

```

DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
139021 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL197365.1 GI:7835515
VERSION AL197365
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143
REFERENCE 3
AUTHORS (bases 1 to 989)
JOURNAL Direct Submission
MEDLINE Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
PUBMED BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1..989
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="139021"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : COAG139AH11LP1-end : T7"

ORIGIN
Query Match      4.5%; Score 68; DB 29; Length 989;
Best Local Similarity 44.9%; Pred. No. 4.6e-05;
Matches 213; Conservative 7; Mismatches 254; Indels 0; Gaps 0;

Qy 171 TAATGCAAAATGATCTGATTCGAATAAAGATGGAGCTTCATTATTTCCTCAAGTAA 230
Db |||
Qy 56 TACVACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 115
Db |||
Qy 231 AGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATTCCTTTTACGCAACTACTACTAGG 290
Db |||
Qy 116 TAATACAAAACACTACWACACTACTACTACTACTACTACTACTACTACTACTACTAA 175
Db |||
Qy 291 AGGAGAACATATGACTACTAAAAGATACGGTCAGCAAAATGAGATTCGACATGGGC 350
Db |||
Qy 176 TACTACTACTAAATCTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 235
Db |||
Qy 351 TAAATCTATATCAGTAAATCTACTACACCAAGTTTCACCTGTTTACTGAGTCAAAATGATGG 410
Db |||
Qy 236 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTATAC 295
Qy 411 TACTGAGTTATTAAGTTTCCCAATATGGAATATTAATATGTTTCTAGCACTGTTAATAA 470
Db |||
Qy 296 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 355

```

```

Qy 471 TGGAGCTGTAATATGTTTACATCTGTAACTCCAAATGCTACTATTTCATGAAAGATAC 530
Db |||
Qy 356 TACTGCTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 415
Db |||
Qy 531 TGATGCGACATGGGAGATGGTGTGGAATAAAGCTGTAGATCAAAAACGTAACGTAACGTTGG 590
Db |||
Qy 416 TACVACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTGG 475
Qy 591 TGATACAGTCAATATATCTACTATTACTTATTAAGAATGTCAGTCAATATATCATCGTAC 644
Db |||
Qy 476 TGGTCTCTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTCC 529

RESULT 6
AQ946120/c
LOCUS AQ946120
DEFINITION Sheared DNA-46J23-TR Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-46J23, genomic survey sequence.
ACCESSION AQ946120
VERSION AQ946120.1 GI:6769385
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 641)
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leech,V., de Jong,P., Ollu,E., Melville,S.,
Donelson,J., Fraser,C. and Adams,M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Other GSSs: Sheared DNA-46J23.TP
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tdb/mdb/tbdb/.
Seq primer: M13-Reverse
Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..641
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-46J23"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site 1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Borell, Oxford University
Press, 1999)."

ORIGIN
Query Match      4.5%; Score 67.8; DB 28; Length 641;
Best Local Similarity 45.3%; Pred. NO. 5e-05;
Matches 246; Conservative 0; Mismatches 297; Indels 0; Gaps 0;

Qy 366 TAATACTACCAAGTTTCCTGTTTACTGAGTCAAAATATGATGCTACTGAGGTATTAA 425

```


The BACs can be obtained from <http://www.incyte.com>.

```

FEATURES
  source
    Location/Qualifiers
      1..501
        /organism="Takifugu rubripes"
        /mol_type="genomic DNA"
        /db_xref="taxon:31033"
        /clone="264E22cA9"
        /clone_lib="BAC 264E22"

ORIGIN
  Query Match      4.4%; Score 66.4; DB 29; Length 501;
  Best Local Similarity 48.9%; Pred. No. 0.0001;
  Matches 178; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 171 TAATGCAATGATGATCTGATTCGAAATAGAGTGGAGCTTCTTATTTAATTCCTCAAGGTAA 230
Db 101 TACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 160
QY 231 AGAGCTGAGTATTAAGCTTCACATGATTTTAAATTCCTTTTAAACGACAACTACTAATGG 290
Db 161 TACTACTACTGCTACTACTACTGCTGCTTCTACTACTACTACTACTACTACTACTACTACTAC 220
QY 291 AGGAGAACATATGATTAACATTAAGAGATCTGCGTCAGCAATGAGATTGGACATGGC 350
Db 221 TAGTACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTGC 280
QY 351 TAAATCTATATCAGTATATACACACAGTTTCACACTGTTTACTGAGTCAATATGATGG 410
Db 281 TACTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 340
QY 411 TACTGAGTTATTAATGTTTCCCAATATGGATATATATATGTTTCTAGCAGTGTATAA 470
Db 341 TACTACTACTACTACTACTACTGCTGCTTCTACTACTACTACTACTACTACTACTACTACTAC 400
QY 471 TGGAGCTGTAATTAATGTTTACTATCTGTAATCCAAATGCTACTATTTCATGAAGAATAC 530
Db 401 TACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTGCTACTACTACTACTACTACTAC 460
QY 531 TGAT 534
Db 461 TGCT 464

RESULT 9
BX729701
LOCUS
DEFINITION
  BX729701 XGC-tadpole Silurana tropicalis cDNA clone TtpA075e10 5',
  mRNA sequence.
ACCESSION
  BX729701
KEYWORDS
  EST.
SOURCE
  Silurana tropicalis (western clawed frog)
  ORGANISM
    Silurana tropicalis
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
    Xenopodinae; Silurana.
    1 (bases 1 to 699)
    Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
    Sanger Xenopus tropicalis EST project 2001 (11_2003)
    Unpublished (2003)
    Contact: Croning MDR
    Sanger Institute
    Hinxton, Cambridgeshire, CB10 1SA, UK
    Email: tropesanger.ac.uk
    Sanger Xenopus tropicalis EST project 2001
    TROPICALIS SEQUENCE ID: TtpA075e10.p1kSP6
    Sequencing primer: SP6
    This sequence is from a Xenopus Gene Collection (XGC) library
    constructed by Nigel Garrett.
    cDNA was oligo dt primed from Sug of poly A+ RNA from tadpole
    embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with
    EcoRI at the 5' end and NotI at the 3' end.
    Vector: pCS107; Site_1: EcoRI; Site_2: NotI

FEATURES
  source
    Location/Qualifiers
      1..699
        /organism="Silurana tropicalis"
        /mol_type="mRNA"
        /db_xref="taxon:8364"
        /clone="TtpA075e10"
        /dev_stage="tadpole (stage 35-40)"
        /lab_host="E. coli DH10B"
        /clone_lib="XGC-tadpole"
        /notes="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
        was oligo dt primed from Sug of poly A+ RNA from tadpole
        embryos. EcoRI-NotI cut cDNA was then ligated into pCS107
        with EcoRI at the 5' end and NotI at the 3' end"

ORIGIN
  Query Match      4.4%; Score 66.4; DB 13; Length 699;
  Best Local Similarity 45.6%; Pred. No. 0.00011;
  Matches 272; Conservative 0; Mismatches 321; Indels 3; Gaps 1;

QY 146 ATAAAGTTTTTGATGCGAGAAATAGATAATGCAAAATGATCTGATTCGAATAAAGATGGAG 205
Db 83 ATACTGATCTACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 142
QY 206 CTTCTTATTTAATTCCTCA---AGGTAAAGAGCTGAGTATAAGCTTCAACTGATTTTA 262
Db 143 GTTCTGCTACTGCTACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 202
QY 263 ATTCTCTTTTACGACAACTACTTAATGGAGGGAGAACATATGTAATAAAGATGACTG 322
Db 203 ATACTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 262
QY 323 CGTCAGCAATAGATGCGCATGCGCTAAATCTATATCAGCTAATATACACACAGTTT 382
Db 263 GTACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 322
QY 383 CCACGTCTTACTGAGTCAAAATATGATGGTACTGAGGTATTAATGTTTCCCAATATGGAT 442
Db 323 GTACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 382
QY 443 ATTATATGTTTCTAGCACTGTTAATATGGAGCTGTAATATGTTTCAATCTGTAACATC 502
Db 383 GTACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 442
QY 503 CAAATGCTACTATTTCATGAAAGAAATACTGATGCGACATGCGAGATGCTGCTGGAATAA 562
Db 443 ATACTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 502
QY 563 CTGTAGATCAAAAAACGTACTCGGTTGGTGATACAGTCAAAATATATATTAATTTATAAGA 622
Db 503 GTACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 562
QY 623 ATGCGATCAATTCATGCTGACAGAAAAGTGTATCAATATGTTTAAAGGATACTATGC 682
Db 563 ATACTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 622
QY 683 CATCTGCTTCTCTAGTTCGATTTGAACGAAGGCTCTTATGAAGTAACTATTACTGAT 738
Db 623 GTACTGGTACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAT 678

RESULT 10
BX780846
LOCUS
DEFINITION
  i120h11.g1 WGS-SpicolorF (DH5a methyl filtered) Sorghum bicolor
  genomic clone i120h11, genomic survey sequence.
ACCESSION
  BX780846
VERSION
  BX780846.1 GI:28958388
KEYWORDS
  GSS.
  Sorghum bicolor (sorghum)
  Sorghum bicolor
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

```


SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
1
REFERENCE Roest Crolius.H., Jaillon.O., Dasilva.C., Bouneau.L., Fisher.C.,
AUTHORS Bernot.A., Fizames.C., Wincker.P., Brottier.P., Quetier.F.,
Saurin.W. and Weissenbach.J.
TITLE Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
2
REFERENCE Roest Crolius.H., Jaillon.O., Dasilva.C., Ozouf-Costaz.C.,
AUTHORS Fizames.C., Fischer.C., Bouneau.L., Billault.A., Quetier.F.,
Saurin.W., Bernot.A. and Weissenbach.J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 10899143
PUBMED 10899143
REFERENCE 3 (bases 1 to 735)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source
1..735
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="123M05"
/clone_lib="G"
/note="Genoscope sequence ID : COBG123AG03LP1-end : T7"
ORIGIN
Query Match 4.3%; Score 64.6; DB 29; Length 735;
Best Local Similarity 44.7%; Pred. No. 0.00027;
Matches 193; Conservative 9; Mismatches 230; Indels 0; Gaps 0;
QY 215 TAATTCCTCAGGTAAGAGCTGAGTATTAAGCTTCAAGCTTCAATCTCTTTT 274
DB 597 TRCWRCTACWACTAANNAANGCNGCGCNACNACTACTAATACAACTACTGCTA 538
QY 275 CGACAACTACTAATGAGGAGGAGACATATGTAACATAAAGATACCTCGCTCAGCAATG 334
DB 537 CTACTACTACTGTAAGAACTACTACTGCTACTACTGCTGCTACTACTACTGCTACWA 478
QY 335 AGATTGGCAATGGCTAAATCTATATACGCTATACCACTACCACTGTTCCACTGTTACTG 394
DB 477 TTKWGTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTGCTG 418
QY 395 AGTCAAAATATGAGTACTGAGTTATTAATGTTCCCAATATGATATATATGTTT 454
DB 417 CTACTACTGCTACTACTACTGCTGCTACTACTGCTGCTGCTGCTGCTGCTACTA 358
QY 455 CTAGCACTGTTAATAAGGAGCTGTAATATGTTGTTACATCTGTAACCTCCAAATGCTACTA 514
DB 357 CTACTACTACTACTACTTWTWCTACTGCTACTACTACTACTACTGCTACTACTACTACTA 298
QY 515 TTCAATGAAAAGAACTACTGATCGGACATGCGGAGATGGTGGGAAAACCTGTAGATCAA 574
DB 297 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 238
QY 575 AAAGCTACTCGTGGTGATACAGTCAAAATATACTATTACTTATAAGAAATGCACTCAAT 634

DB 237 CTACTGCTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTA 178
QY 635 ATCATGTTACAG 646
DB 177 CTACTGCTACTG 166
RESULT 14
BZ643398
LOCUS OGAOW577C ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA011J17,
DEFINITION genomic survey sequence.
ACCESSION BZ643398
VERSION BZ643398.1 GI:28104877
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 480)
REFERENCE Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
AUTHORS Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other GSSs: OGAOW57TM
Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source
1..480
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA011J17"
/clone_lib="ZM 0.7 1.5 KB"
/note="Vector: pBSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"
ORIGIN
Query Match 4.2%; Score 62.8; DB 28; Length 480;
Best Local Similarity 48.8%; Pred. No. 0.0007;
Matches 200; Conservative 0; Mismatches 207; Indels 3; Gaps 1;
QY 130 GCAACCTATAAAGCATATAAAGTTTTTGTATGACAGAAATAGATAATGCAATGTATCTGAT 189
DB 17 GCTACTACTACTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 76
QY 190 TCGAATAAGATGGAGCTTCTTATTTAATTCCTCAAG---GTAAGAAGCTGAGTATAAA 246
DB 77 ACTACTACTACTGCTGCTGCTACTACTACTACTGCTGCTGCTACTACTACTGCTGCTACT 136
QY 247 GCTTCAACTGATTTAATTTCTTTTACGACAACTACTAATGAGGAGGAGCAATATGTA 306
DB 137 ACTACTACTACTGCTGCTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 196
QY 307 ACTAAAAAGATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 366
DB 197 ACTGCTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 256
QY 367 AATACTACACCACTTCCACTGTTACTGAGTCAAAATAGTGGTACTGAGGTATTAAT 426
DB 257 ACTACTGCTACTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 316
QY 427 GTTTCCCAATATGATATTTATTTCTAGCACTGTTAATGAGCTGTAATATG 486

Db 317 GCTACTGCTACTGCTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 376

QY 487 GTTACATCTGTAACCCAAATGCTACTATTTCATGAAGAATACTCATGC 536

Db 377 ACTGTCGTACTGCTACTGCTACTGCTAACTGCTACTACTGCTACTAC 426

RESULT 15					
BZ423698					
LOCUS	BZ423698	513 bp	DNA	linear	GSS 10-DEC-2002
DEFINITION	id52c10.g1 WGS-SbicolorF (DH5a methyl filtered) Sorghum bicolor genomic clone id52c10 5', genomic survey sequence.				

ACCESSION	BZ423698	
VERSION	BZ423698.1	GI:26373196
KEYWORDS	GSS.	
SOURCE	Sorghum bicolor	(sorghum)
ORGANISM	Sorghum bicolor	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.	

REFERENCE
AUTHORS
1 (bases 1 to 513)
Rabinovich, P.D., O'Shaughnessy, A.L., Ballja, V., Dedhia, N.,
Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
Zutavern, T., Palmer, L., McCombie, W.R. and Martensen, R.A.
TITLE
JOURNAL
Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
Unpublished (2002)

Captured on 12/02/
 Contact: W. Richard McCombie
 Lita Annenberg Hazen Genome Sequencing Center
 Cold Spring Harbor Laboratory
 PO Box 100, Cold Spring Harbor, NY 11724, USA
 Tel: 516 367 8884

Tel: 516 367 0804
 Fax: 516 367 8874
 Email: mccombie@cshl.org
 Plate: id52 row: c column: 10
 Seq primer: -21M13univRev
 Class: shotgun
 High quality sequence stop: 513

```

FEATURES
  source
    Location/Qualifiers
      1..513
        /organism="Sorghum bicolor"
        /mol_type="genomic DNA"
        /db_xref="taxon:4558"
        /clone_id="d52c10"
        /lab_host="DH5a"

```

```

/clone_lib="WGS-SbicolorF (DH5a methyl filtered)"
digested="Xba I; Site 2: Xba I; The vector was
ligated with Xba I and one nucleotide was added by fill in
in the recessive 3' end. The genomic DNA was nebulized,
end repaired, adaptor ligated and size fractionated using
sephadex. The resulting fragments were between 0.8 and 3
kb and were cloned into the vector (-x/y reads in M13mp19,
b/g reads in pUC19). The same ligation was transformed
into DH5a."

```

ORIGIN

Query Match	4.1%	Score 62.6;	DB 28;	Length 513;
Best Local Similarity	46.3%;	Prod. No. 0.00078;		
Matches 206;	Conservative	0;	Mismatches 239;	Indels 0; Gaps 0;
171	TAATGCAAAATGATCTGATTCGAATAAGATGGAGCTTTCTTTAATTCTCTCAAGTAA	230		
65	TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC	124		
231	AGAGCTTGAGTATAAAGCTTCAACTGATTTTAATTTCTCTTTTACGACAACTACTAATGG	290		
125	TAC	184		
291	AGGAGAGCAATATGACTATAAAGATGATCGGTGACCAAAATGAGATTGGCATGGGC	350		
185	TAGTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC	244		
351	TAAATCTATATGAGCTAATACTACACCAAGTTTCCACTGTTACTGAGTCAAAATAATGATGG	410		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2004, 22:30:05 ; Search time 111.254 Seconds
(without alignments)
7527.109 Million cell updates/sec

Title: US-10-009-254-1

Perfect score: 1509

Sequence: 1 atgaaagaaatgattca.....gtcgtcgtttaagttcttaa 1509

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64.8	4.3	1980	4	US-09-134-000C-2719
2	64.8	4.3	1983	4	US-09-071-035-217
3	56.2	3.7	1575	4	US-09-543-581A-615
4	50.4	3.3	14066	4	US-09-601-198-56
5	49.8	3.3	1687	4	US-09-071-035-219
6	48.4	3.2	3666	2	US-08-682-517-13
7	48.4	3.2	3666	2	US-08-682-517-14
8	48.4	3.2	4197	2	US-08-682-517-7
9	48.4	3.2	4197	2	US-08-682-517-8
10	47.8	3.2	740	3	US-09-451-117-1
11	47.8	3.2	740	4	US-09-888-655-1
12	47.4	3.1	1037	4	US-09-181-585-3
13	47.4	3.1	1159	4	US-09-181-585-1
14	47.4	3.1	1471	4	US-09-181-585-2
15	47.4	3.1	3945	4	US-09-200-6508-6
16	47.4	3.1	4358	4	US-08-956-171E-454
17	47.2	3.1	7218	1	US-08-232-463-14
18	46.8	3.1	1948	1	US-07-849-438-1
19	46.8	3.1	4590	4	US-09-134-001C-1108
20	45.6	3.0	4376	1	US-08-119-125A-1
21	45.6	3.0	6744	1	US-08-119-125A-2
22	45	3.0	832	4	US-09-621-976-2813
23	44.6	3.0	11679	4	US-09-328-352-1377
24	44.2	2.9	606	4	US-09-601-198-166
25	44.2	2.9	30549	4	US-09-134-001C-322
26	44	2.9	1716	4	US-09-134-001C-1028
27	43.4	2.9	2763	4	US-09-463-402-5

28	43.4	2.9	2765	4	US-09-889-572-13	Sequence 3, Appli
29	43.4	2.9	4085	4	US-09-463-402-13	Sequence 13, Appl
30	43.2	2.9	5181	1	US-08-257-073-10	Sequence 10, Appl
31	43	2.8	2001	4	US-09-528-784A-84	Sequence 84, Appl
32	43	2.8	2001	4	US-09-569-098A-84	Sequence 84, Appl
33	43	2.8	3057	4	US-09-601-198-55	Sequence 55, Appl
34	43	2.8	3402	4	US-09-528-784A-86	Sequence 86, Appl
35	43	2.8	3402	4	US-09-569-098A-86	Sequence 86, Appl
36	42.8	2.8	966	4	US-09-134-001C-1789	Sequence 1789, Ap
37	42.6	2.8	168575	4	US-09-426-290-1	Sequence 1, Appli
38	42.2	2.8	810	4	US-09-601-198-31	Sequence 31, Appl
39	42.2	2.8	3292	1	US-07-814-964-12	Sequence 12, Appl
40	42.2	2.8	3292	1	US-08-258-442-12	Sequence 12, Appl
41	42.2	2.8	3292	1	US-08-328-809-7	Sequence 7, Appli
42	42.2	2.8	3292	4	US-08-866-840-7	Sequence 7, Appli
43	42.2	2.8	3292	5	PCT-US92-11107-12	Sequence 12, Appl
44	42.2	2.8	4185	4	US-09-417-485D-7	Sequence 7, Appli
45	42.2	2.8	10640	4	US-09-417-485D-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-134-000C-2719

; Sequence 2719, Application US/09134000C

; Patent No. 6617156

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/134,000C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/055,778

; PRIOR FILING DATE: 1997-08-15

; NUMBER OF SEQ ID NOS: 6812

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 2719

; LENGTH: 1980

; TYPE: DNA

; ORGANISM: Enterococcus faecalis

; US-09-134-000C-2719

Query Match	4.3%;	Score 64.8;	DB 4;	Length 1980;
Best Local Similarity	52.1%;	Pred. No. 9.4e-07;		
Matches 172;	Conservative	0;	Mismatches 152;	Indels 6; Gaps 1;
Qy	1180	ACAGAAGCTTAATCGACAGAAATATACAA	CAGGAGCAGATGGTATATATTACCATTCAGGC	1239
Db	1657	ACAAAAGCTGAAGCAACTACTTTTACAA	CACGGCTGATGGATTAGTTGATATCACAGGG	1716
Qy	1240	TTGAAAGAAGGTACATACTACTTAGTTG	AGAAAAGGCTCCCTTAGGTTCACAAATTTGTTA	1299
Db	1717	CTTAAATACGGTACCTTATTTAGAGA	AAACTGTAGTCTCTCGATGATTATGTCCTTGTTA	1776
Qy	1300	GATAACTCTCAGAAGGTTATTTTAGGAG	ATGGAGCTGAGGCCACTGATACGACTAATTACAGATAAC	1359
Db	1777	ACAAATCGGATTGAATTTGGTCAATGA	ACAAATCATATGGCAAC-----CAGAAAAC	1830
Qy	1360	CTTTTAGTTAAACCAACTGTTGAAAATA	TACAAAGGTACTGAGTTGCCCTTCAACAGGTGGT	1419
Db	1831	CTAGTTTTCACGAAAAGTACCAACAA	ACAAACAAAGGTACCTTACCTTCAACAGGTGGC	1890
Qy	1420	ATTGGTACAACAATTTTCTACATTAT	PAGGTGCAATTTTAGTATATAGGAGCAGGATTCGTG	1479
Db	1891	AAAGGAATCTACGTTTACTTTAGGA	AGTGGCGAGCTTGTCTACTTATTCGAGAGTCTAC	1950

APPLICANT: Chen, Ellison Y.
APPLICANT: Glass, Jennifer S.
APPLICANT: Glass, John I.
APPLICANT: Heiner, Cheryl R.
APPLICANT: Lefkowitz, Elliot
TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
FILE REFERENCE: UAB-13452/22
CURRENT APPLICATION NUMBER: US/09/601,198
CURRENT FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/073,189
PRIOR FILING DATE: 1998-01-30
NUMBER OF SEQ ID NOS: 181
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 56
LENGTH: 14066
TYPE: DNA
ORGANISM: Ureaplasma urealyticum
US-09-601-198-56

Query Match 3.3%; Score 50.4; DB 4; Length 14066;
Best Local Similarity 42.8%; Pred. No. 0.007;
Matches 313; Conservative 0; Mismatches 416; Indels 3; Gaps 1;

QY 435 ATATGATATTATTTCTTCTAGCACTGTTAATATGGAGCTGTAATATGTTTACATC 494
DB 5284 AAATGTTAGTTCTTAATGATTTCTAATAATCTACAACAACAGTAAATATGATTTGA 5225

QY 495 TGTAACTCCAAATGCTACTATTCTATGAAAGAAATCTGATGCGACATGGGAGATGGTGG 554
DB 5224 TGGAAATCAAGCACTTGATTAATAAATAATCAATTTGGTATATATCTCAATGACGG 5165

QY 555 TGGAAAACTGTAGATCAAAAAACGTAAGTCTGGTGGTATGATGATCAATATATCTATTAC 614
DB 5164 TGAAGAAATTTAAGTGATCAAAAAACATTCCTTGGGCTAATAACCACTATAGTTTGA 5105

QY 615 TTATAGAATGCAATCAATATCATGTCAGAGAAAGTGTATCAATATGTTTAAAGGA 674
DB 5104 ATTGTCTAATTTAAACATCAATCGTAAATATACACTAAAAGAGTAAAGTAAATGA 5045

QY 675 TACTATGCCATCTGCTCTCTAGTCTGATTTGAAGAGAGGCTCTTATGAAGTAACTATTAC 734
DB 5044 TGATATAAACAATCAATAATTTTCAATTTAAATAATGATAGCTGATTTGATTTGT 4985

QY 735 TGATGATCAGGAATATTAACAATCTAATCAAGTTTCGAAAAAGCACTGGGAAGTA 794
DB 4984 TAATAAAACACAATCGATTAAGTTCAATTAATGAGCCAAATGCAAGAGCGAAAAA 4925

QY 795 TAACCTGTTAGAGAAAAATTAATTTTCAAGTACTATTCCTGGGCGAGTACCATAAC 854
DB 4924 TAATTTTACAAATCAACCAATTAAGATTTATATTAATGATCTCTGATAATGTTTAAAGTAA 4865

QY 855 TCAACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTTTATAAGGGAATAATAC 914
DB 4864 TGAAGAAGAGCAATCAATTAATTTATGTTGAGAT---TTATCTGTAAGCAAAAGTTAA 4808

QY 915 AATCAGAGTCATTTATACAGAGATTTTAAAGAGTGGAGCTTAAACACAGGTTTCAGCTGATT 974
DB 4807 AATTGAGGAGTCAAAAATTTTAAAGAACCAATCAATCAATTAATGATTAATGTTTAAAGTAA 4748

QY 975 ACAGAAAAATCAAAACATTCGAGCAATCAACCCCAATAGCAATGATGACCCAGGTCA 1034
DB 4747 TACTATTATTAACAGTATTACTTTTATTAATAACCAATAAAGCTGCTACGAATATTGG 4688

QY 1035 AAAAGTAACAGTACGAGGATGCTCAATTAATTAATAAATAATGATGGTTCCCAAAAGC 1094
DB 4687 AATAGATAATAGCAATAAATAATATACAAATCAACAAATATAATCAATTAATAAATTGA 4628

QY 1095 TTCAATCAAGGCTGCTATATTTGTTTTTAAAGATGCTTACGGGTCAATTTCTAAACTTTAA 1154
DB 4627 TAATAATTTAGTGTCTGCTCTGTTCAAGTGATCACTAATAACGTTACTTACTTTGA 4568

QY 1155 CGATACAAAATAA 1166

DB 4567 ATATAATAATAA 4556

RESULT 5
US-09-071-035-219
Sequence 219, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 219:
SEQUENCE CHARACTERISTICS:
LENGTH: 1687 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-219

Query Match 3.3%; Score 49.8; DB 4; Length 1687;
Best Local Similarity 62.4%; Pred. No. 0.0047;
Matches 78; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1180 ACAGAAGCTAATGCAACAGAAATATACACAGGAGCAGATGGTATAATTACATTACAGGC 1239
DB 1466 ACAAAAGCTGAAGCAACTACTTTTACAAACAAACGCTGATGATTTGATATACACAGG 1525

QY 1240 TTCAAGAGAGGTACATACATCTAGTTGAGAAAAAGCTCCCTTAGGTTACAATTTGTTA 1299
DB 1526 CTTAAATACGGTACCTATTATTATTAGAAGAAACTGTAGTCTCTGATGATTATGTTGTTA 1585

QY 1300 GATAA 1304
DB 1586 ACAAA 1590

RESULT 6
US-08-682-517-13
Sequence 13, Application US/08682517
Patent No. 5874267
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Expression of surface layer proteins
NUMBER OF SEQUENCES: 25
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/682,517
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 3666 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-682-517-13

Query Match 3.2%; Score 48.4; DB 2; Length 3666;
Best Local Similarity 44.5%; Pred. No. 0.014;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 302 ATGTAAGTCAACAACTGTTGATACGCACTGTTTCATTAAGAGATAGTGCAAAATCTATAT 361
DB 2768 ATGTAAGTCAACAACTGTTGATACGCACTGTTTCATTAAGAGATAGTGCAAAATCTATAT 2827

QY 362 CAGCTAATACACAGAGTTCCACTGTTACTGAGTCAAAATAGTGGTACTGAGGTTA 421
DB 2828 CATTATCTCTTACATTAAGTTGAAACTGGTCTAATACAGGTGATTTTGCTACAACTGTTT 2887

QY 422 TTAATGTTTCCCAATATGGATATTTATTTAGTTTCTAGCACTGTTAATTAATGGAGCTGTA 481
DB 2888 AAGCTGTGATCAATTAATCTTTAACTGCTGATCAATTAAGAGATAGTGCAAAATCTATAT 2947

QY 482 TTATGTTTACATCTGTAACCTCCAAATGCTACTATTATTCATGAAAGAAATGCTGCGACAT 541
DB 2948 AAAATGCTGAGGTGTTGCTGAAATATTTACTGCTAGCGTAAACATTAAGAGAACTACTG 3007

QY 542 GGGAGATGTTGGTGAAGAACTGTAGATCAAAAAGCTACTCGGTTGGTGATACAGTCA 601
DB 3008 GAGCAATTAATCTCTGATACATTTACACAAAGGTGATTTACCATCAGCAGCTACAGAGCTG 3067

QY 602 AATATACTATTACTTAAAGATGCACTCAATATCATGTCAGAAAAAGTGTATCAAT 661
DB 3068 AATATACTTCTAATCAATGCTGCGATTTATACATTTGCAACAGGTGAAGGATTCATT 3127

QY 662 ATGTTAAGAGTACTATGCCATCTGCTCTCTGATGTTGATTTGAACGAGGCTCTTATG 721
DB 3128 TAAATATTGATGATGCTGCTGCTCAAGTAAATTAATAGCAGGTAAAAAGGTGCACAAG 3187

QY 722 AAGTAATATTACT 735
DB 3188 GTGTAGCTGATGCT 3201

RESULT 7
US-08-682-517-14
Sequence 14, Application US/08682517
Patent No. 5874267
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Expression of surface layer proteins
NUMBER OF SEQUENCES: 25
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/682,517
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 3666 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3666
US-08-682-517-14

Query Match 3.2%; Score 48.4; DB 2; Length 3666;
Best Local Similarity 44.5%; Pred. No. 0.014;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 302 ATGTAAGTCAACAACTGTTGATACGCACTGTTTCATTAAGAGATAGTGCAAAATCTATAT 361
DB 2768 ATGTAAGTCAACAACTGTTGATACGCACTGTTTCATTAAGAGATAGTGCAAAATCTATAT 2827

QY 362 CAGCTAATACACAGAGTTCCACTGTTACTGAGTCAAAATAGTGGTACTGAGGTTA 421
DB 2828 CATTATCTCTTACATTAAGTTGAAACTGGTCTAATACAGGTGATTTTGCTACAACTGTTT 2887

QY 422 TTAATGTTTCCCAATATGGATATTTATTTAGTTTCTAGCACTGTTAATTAATGGAGCTGTA 481
DB 2888 AAGCTGTGATCAATTAATCTTTAACTGCTGATCAATTAAGAGATAGTGCAAAATCTATAT 2947

QY 482 TTATGTTTACATCTGTAACCTCCAAATGCTACTATTATTCATGAAAGAAATGCTGCGACAT 541
DB 2948 AAAATGCTGAGGTGTTGCTGAAATATTTACTGCTAGCGTAAACATTAAGAGAACTACTG 3007

QY 542 GGGAGATGTTGGTGAAGAACTGTAGATCAAAAAGCTACTCGGTTGGTGATACAGTCA 601
DB 3008 GAGCAATTAATCTCTGATACATTTACACAAAGGTGATTTACCATCAGCAGCTACAGAGCTG 3067

QY 602 AATATACTATTACTTAAAGATGCACTCAATATCATGTCAGAAAAAGTGTATCAAT 661
DB 3068 AATATACTTCTAATCAATGCTGCGATTTATACATTTGCAACAGGTGAAGGATTCATT 3127

QY 662 ATGTTAAGAGTACTATGCCATCTGCTCTCTGATGTTGATTTGAACGAGGCTCTTATG 721
DB 3128 TAAATATTGATGATGCTGCTGCTCAAGTAAATTAATAGCAGGTAAAAAGGTGCACAAG 3187

QY 722 AAGTAATATTACT 735
DB 3188 GTGTAGCTGATGCT 3201

RESULT 8
US-08-682-517-7
Sequence 7, Application US/08682517
Patent No. 5874267
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Expression of surface layer proteins
NUMBER OF SEQUENCES: 25
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/682,517
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4197 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-682-517-7

Query Match 3.2%; Score 48.4; DB 2; Length 4197;

Best Local Similarity 44.5%; Pred. No. 0.014; Mismatches 241; Indels 0; Gaps 0;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;
QY 302 ATGTAACATAAAAGATGCTGCGTCAGCAAAATGAGATTGGCATGGCTTAAATCTATAT 361
DB 2952 ATGTAAGTGAACAACATGTTGATGCTGCAACTGTTTCATTAAAGATAGTGCAATAATT 3011
QY 362 CAGCTAATATACACCAAGTTTCCAGTGTCTGAGTCAAAATGATGCTGAGGTGA 421
DB 3012 CATTATCTCTTACATTAGTTGAAACTGGTGTAAATACAGGTGATTGCTACCAACTGTC 3071
QY 422 TTAATGTTTCCCAATATGATATATTATGTTTCTAGCACTGTTAATAATGGAGCTGTA 481
DB 3072 AAGCTGGTACATTATCTTTAACTGCTGGTATCAATTAACAGTTACTTATGCGAGTCTA 3131
QY 482 TTATGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAAAGAAATGATGCGACAT 541
DB 3132 AAAATGCTGAGGTGTTGCTGAAAAATATTACTGCTAGCTGAACATTAAGAAAACTACTG 3191
QY 542 GGGAGATGCTGGTGGAAAACTGTAGATCAAAAACGTACTCGGTTGGTGATACAGTCA 601
DB 3192 GAGCAATTTACTTCTGATACATTTACACAAGGTGATTACCATCAGCAGCTACAGCAGCTG 3251
QY 602 AATATATCTTACTTATTAAGATGCAATGCAATTAATCATGTCAGAAAAAGTGTATCAAT 661
DB 3252 AATATATCTTCTAAATCAATTTGCTGCAATTAATCAATTTGCAACAGGTGAAGGATTCATT 3311
QY 662 ATGTTTAAAGGATCTATGCCATCTGCTTCTGCTAGTTGATTGAAACGAAGGGTCTTTATG 721
DB 3312 TAAATATTGATATGCTGGTGCTCAAGTAATTAACCTAGCAGGTAAAAAGGTGCACAAG 3371
QY 722 AAGTAATCTTACT 735
DB 3372 GTGTAGCTGATGCT 3385

RESULT 9

US-08-682-517-8
; Sequence 8, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus sphaericus
; INDIVIDUAL ISOLATE: P-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..3850
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 185..3850
; FEATURE:
; NAME/KEY: sig peptide
; LOCATION: 95..184
; US-08-682-517-8

Query Match 3.2%; Score 48.4; DB 2; Length 4197;
Best Local Similarity 44.5%; Pred. No. 0.014; Mismatches 241; Indels 0; Gaps 0;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;
QY 302 ATGTAACATAAAAGATGCTGCGTCAGCAAAATGAGATTGGCATGGCTTAAATCTATAT 361
DB 2952 ATGTAAGTGAACAACATGTTGATGCTGCAACTGTTTCATTAAAGATAGTGCAATAATT 3011
QY 362 CAGCTAATATACACCAAGTTTCCAGTGTCTGAGTCAAAATGATGCTGAGGTGA 421
DB 3012 CATTATCTCTTACATTAGTTGAAACTGGTGTAAATACAGGTGATTGCTACCAACTGTC 3071
QY 422 TTAATGTTTCCCAATATGATATATTATGTTTCTAGCACTGTTAATAATGGAGCTGTA 481
DB 3072 AAGCTGGTACATTATCTTTAACTGCTGGTATCAATTAACAGTTACTTATGCGAGTCTA 3131
QY 482 TTATGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAAAGAAATGATGCGACAT 541
DB 3132 AAAATGCTGAGGTGTTGCTGAAAAATATTACTGCTAGCTGAACATTAAGAAAACTACTG 3191
QY 542 GGGAGATGCTGGTGGAAAACTGTAGATCAAAAACGTACTCGGTTGGTGATACAGTCA 601
DB 3192 GAGCAATTTACTTCTGATACATTTACACAAGGTGATTACCATCAGCAGCTACAGCAGCTG 3251
QY 602 AATATATCTTACTTATTAAGATGCAATGCAATTAATCATGTCAGAAAAAGTGTATCAAT 661
DB 3252 AATATATCTTCTAAATCAATTTGCTGCAATTAATCAATTTGCAACAGGTGAAGGATTCATT 3311
QY 662 ATGTTTAAAGGATCTATGCCATCTGCTTCTGCTAGTTGATTGAAACGAAGGGTCTTTATG 721
DB 3312 TAAATATTGATATGCTGGTGCTCAAGTAATTAACCTAGCAGGTAAAAAGGTGCACAAG 3371
QY 722 AAGTAATCTTACT 735
DB 3372 GTGTAGCTGATGCT 3385

RESULT 10

US-09-451-117-1
; Sequence 1, Application US/09451117
; Patent No. 6277973
; GENERAL INFORMATION:
; APPLICANT: Jenkins, Mark C.
; APPLICANT: Fayer, Ronald
; APPLICANT: Trout, James
; TITLE OF INVENTION: Cloning and Expression of a DNA Sequence Encoding A 41
; Patent No. 6277973
; TITLE OF INVENTION: kDa Cryptosporidium parvum Oocyst Wall Protein
; FILE REFERENCE: 0046.99
; CURRENT APPLICATION NUMBER: US/09/451,117
; CURRENT FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
; US-09-451-117-1

Query Match 3.2%; Score 47.8; DB 3; Length 740;
Best Local Similarity 46.7%; Pred. No. 0.011; Mismatches 172; Indels 0; Gaps 0;
Matches 151; Conservative 0; Mismatches 172; Indels 0; Gaps 0;
QY 217 ATTCTCAAGGTAAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAATTTCTTTTACG 276
DB 370 ACTACTCAATCAATGGAAGTACTGATATAATATTGGTAAATATGAGTGGTATTATTACA 429
QY 277 ACAACTACTAATGGAGGGAACAATATGTAACATAAAAGAACTACTGCGTCAGCAAAATGAG 336
DB 430 TCTAGTGGTATCTTAGCTTTACTTAATACTTAATCTTAATGTTAATAATAGTAATAGT 489
QY 337 ATTGGCAGATGGGCTAAATCTATATACAGTAATTAACACTACACAGTTTCCACTGTTACTGAG 396

Db 490 AATATTGGATCAGAAATTTTATACAGTTGGTACTTGTCTTCTACTAGTATTGGTAAT 549
QY 397 TCAAAATGAATGAGTACTGAGGTTATTAATGTTTCCCAATATGGATATATATGTTTCT 456
Db 550 AGTAATGAGTGTCTTTTACTGCTATTCATCCATAATAAACAAPAGCAATAATTAAT 609
QY 457 AGCACTGTTAATPAATGAGGCTGTAATTTATGTTTACATCTGTAACCTCCAAATGCTACTATT 516
Db 610 AATAATAATAATAATAAGTAATACCACTCTTACTACTGTTGCTACTAATGCTAATATT 669
QY 517 CATGAAGAATACTGATCGGAC 539
Db 670 ACTACTAATACTACTAATACTACTAC 692

RESULT 11
US-09-888-655-1
; Sequence 1, Application US/09888655
; Patent No. 6521229
; GENERAL INFORMATION:
; APPLICANT: Jenkins, Mark C.
; APPLICANT: Fayer, Ronald
; APPLICANT: Trout, James
; TITLE OF INVENTION: Cloning and Expression of a DNA Sequence Encoding A 41
; Patent No. 6521229
; TITLE OF INVENTION: kDa Cryptosporidium parvum Oocyst Wall Protein
; FILE REFERENCE: 0046.99
; CURRENT APPLICATION NUMBER: US/09/888,655
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-09-888-655-1

Query Match 3.2%; Score 47.8; DB 4; Length 740;
Best Local Similarity 46.7%; Pred. No. 0.011;
Matches 151; Conservative 0; Mismatches 172; Indels 0; Gaps 0;
QY 217 ATTCTCAAGGTAAAGAGCTGAGTATAAAGCTTCAACTGATTTTAATTTCTTTTACG 276
Db 370 ACTACTCAATCAATGGAAGTTACTGATACTAATAATGTTAATGATGATGATGATGAT 429
QY 277 ACACTACTAATGAGGAGAGACATATGTAATAAAGAGTACTGCGTCAGCAAAATGAG 336
Db 430 TCTAGTGTGATTTCTATAGCTGTTACTAATAATCTTAATGGTAATAATAATAGTAAT 489
QY 337 ATTCCGACATGGGCTAAATCTATATCAGCTAATACTACACAGTTTCCACTGTTACTGAG 396
Db 490 AATAATGGATCAGAAATTTTATACCAAGTTGGTACTTGTCTTCTACTAGTATTGGTAAT 549
QY 397 TCAATAATGATGCTACTGAGGTTAATAATGTTTCCCAATATGGATATTAATGTTTCT 456
Db 550 AGTAATGCTGTGCTTTTACTGCTATTCATCCATAATAAACAATAGCAATAATAATTAAT 609
QY 457 AGCACTGTTAATGAGGCTGTAATTTATGTTTACATCTGTAACCTCCAAATGCTACTATT 516
Db 610 AATAATAATAATAATAAGTAATACCACTCTTACTACTGTTGCTACTAATGCTAATATT 669
QY 517 CATGAAGAATACTGATCGGAC 539
Db 670 ACTACTAATACTACTAATACTACTAC 692

RESULT 12
US-09-181-585-3
; Sequence 3, Application US/09181585
; Patent No. 6524791
; GENERAL INFORMATION:
; APPLICANT: Ranum, Laura P.W.

; APPLICANT: Koob, Michael
; TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 1100900101
; CURRENT APPLICATION NUMBER: US/09/181,585
; CURRENT FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1037
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: cDNA
; OTHER INFORMATION: comprising exons E, C, and A
US-09-181-585-3

Query Match 3.1%; Score 47.4; DB 4; Length 1037;
Best Local Similarity 47.2%; Pred. No. 0.016;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
QY 218 TTCCTCAAGGTAAAGAGCTGAGTATAAAGCTTCAACTGATTTTAATTTCTTTTACGA 277
Db 611 TCCTTCATGTTAGAAAACCTGGCTTTTACTACTACTACTACTACTACTACTACTACTA 670
QY 278 CAACTACTAATGAGGAGAGAACATATGTAATAAAGAGTACTGCGTCAGCAAAATGAGA 337
Db 671 CTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 730
QY 338 TTGCGACATGGGCTAAATCTATATACAGCTAATACTACACAGTTTCCACTGTTACTGAGT 397
Db 731 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 790
QY 398 CAAATAATGATGCTACTGAGGTTTAAATGTTTCCCAATATGGATATTAATGTTTCTA 457
Db 791 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 850
QY 458 GCACTGTTAATAATGAGGCTGTAATTTATGTTTACATCTGTAACCTCCAAATGCTACTATT 517
Db 851 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 910
QY 518 ATGAA 522
Db 911 CTGCA 915

RESULT 13
US-09-181-585-1
; Sequence 1, Application US/09181585
; Patent No. 6524791
; GENERAL INFORMATION:
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Koob, Michael
; TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 1100900101
; CURRENT APPLICATION NUMBER: US/09/181,585
; CURRENT FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1159
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-181-585-1

Query Match 3.1%; Score 47.4; DB 4; Length 1159;
Best Local Similarity 47.2%; Pred. No. 0.016;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
QY 218 TTCCTCAAGGTAAAGAGCTGAGTATAAAGCTTCAACTGATTTTAATTTCTTTTACGA 277
Db 423 TCCTTCATGTTAGAAAACCTGGCTTTTACTACTACTACTACTACTACTACTACTACTA 482
QY 278 CAACTACTAATGAGGAGAGAACATATGTAATAAAGAGTACTGCGTCAGCAAAATGAGA 337

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 01:51:02 ; Search time 730.222 Seconds
(without alignments)
10397.614 Million cell updates/sec

Title: us-10-009-254-1
Perfect score: 1509
Sequence: 1 atgaaaagaaatgattca.....gtcgctgttaagttcttaa 1509

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters: 6608766

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	1509	100.0	1509	12	US-10-333-002-15
2	64.8	4.3	1983	13	US-10-206-576-217
3	64.8	4.3	2199	12	US-10-333-002-28
4	64.8	4.3	15614	9	US-09-070-927A-45
5	63.2	4.2	1887	17	US-10-661-809-12
6	59.2	3.9	1131	13	US-10-282-122A-15988
7	53	3.5	4985	15	US-10-094-240-10
8	53	3.5	4985	15	US-10-056-405-10
9	50.4	3.3	14066	15	US-10-349-680-149
10	50.4	3.3	14067	13	US-10-282-122A-40681
11	50	3.3	3996	15	US-10-087-464-42
12	49.8	3.3	1687	13	US-10-206-576-219
13	48.6	3.2	4997	13	US-10-282-122A-35506
14	48.4	3.2	3666	9	US-09-137-531-13

15	48.4	3.2	3666	9	US-09-137-531-14	Sequence 14, Appl
16	48.4	3.2	4197	9	US-09-137-531-7	Sequence 7, Appl
17	48.4	3.2	4197	9	US-09-137-531-8	Sequence 8, Appl
18	47.4	3.1	1037	16	US-10-373-667-3	Sequence 3, Appl
19	47.4	3.1	1159	16	US-10-373-667-1	Sequence 1, Appl
20	47.4	3.1	1471	16	US-10-373-667-2	Sequence 2, Appl
21	47.4	3.1	4047	9	US-09-815-242-4843	Sequence 4843, Ap
22	47.4	3.1	4050	9	US-09-815-242-9039	Sequence 9039, Ap
23	47.4	3.1	4358	8	US-08-781-986A-454	Sequence 454, App
24	47.4	3.1	4358	13	US-10-329-624-454	Sequence 454, App
25	47.2	3.1	5314	15	US-10-155-533-1	Sequence 1, Appl
26	46.8	3.1	2801	13	US-10-282-122A-34844	Sequence 34844, A
27	46.8	3.1	7047	15	US-10-240-453-260	Sequence 260, App
28	46.2	3.1	4158	13	US-10-282-122A-8140	Sequence 8140, Ap
29	45.8	3.0	684707	16	US-10-398-221-9	Sequence 9, Appl
30	45.8	3.0	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
31	45.6	3.0	2017	15	US-10-155-533-3	Sequence 3, Appl
32	45.6	3.0	6397	13	US-10-221-714A-181	Sequence 181, App
33	45.6	3.0	6397	15	US-10-239-676-107	Sequence 107, App
34	45.6	3.0	6397	15	US-10-311-455-1319	Sequence 1319, Ap
35	45.6	3.0	6397	15	US-10-240-453-119	Sequence 119, App
36	45.4	3.0	2217	13	US-10-282-122A-17439	Sequence 17439, A
37	45.2	3.0	3931	15	US-10-006-780-1	Sequence 1, Appl
38	45	3.0	12237	15	US-10-311-455-2331	Sequence 2331, Ap
39	44.8	3.0	573	15	US-10-029-386-25433	Sequence 25433, A
40	44.8	3.0	5358	17	US-10-637-544-19	Sequence 19, Appl
41	44.8	3.0	16287	15	US-10-311-455-645	Sequence 645, App
42	44.6	3.0	954	17	US-10-451-467A-537	Sequence 537, App
43	44.4	2.9	7446	13	US-10-282-122A-7578	Sequence 7578, Ap
44	44.2	2.9	486	13	US-10-465-217-13	Sequence 13, Appl
45	44.2	2.9	606	15	US-10-349-680-85	Sequence 85, Appl

ALIGNMENTS

RESULT 1

US-10-333-002-15
; Sequence 15, Application US/10333002
; Publication No. US20040071729A1

; GENERAL INFORMATION:

; APPLICANT: Adderson, Elisabeth

; APPLICANT: Bohnsack, John

; TITLE OF INVENTION: Group B Streptococcus Polypeptides Nucleic Acids and Therapeutic

; TITLE OF INVENTION: Compositions and Vaccines Thereof

; FILE REFERENCE: 2511-1-001 (SJ-0039)

; CURRENT APPLICATION NUMBER: US/10/333,002

; CURRENT FILING DATE: 2003-07-08

; PRIOR APPLICATION NUMBER: PCT/US01/24795

; PRIOR FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 09/634,341

; PRIOR FILING DATE: 2000-08-08

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 15

; LENGTH: 1509

; TYPE: DNA

; ORGANISM: Streptococcus agalactiae

US-10-333-002-15

Query Match 100.0%; Score 1509; DB 12; Length 1509;
Best Local Similarity 100.0%; Pred. No. 6.8e-314;
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATCAAAAGAAATGATTCAATCGCTGTAGTCGCGAGTTTAGCATTGGTATGCGCTGTA 60

Db 1 ATCAAAAGAAATGATTCAATCGCTGTAGTCGCGAGTTTAGCATTGGTATGCGCTGTA 60

Qy 61 TCACCAAGTACGCGCGATAGCTTTTGGCGCTGAGACAGGGAACAATTACAGTTCAAGTACT 120

Db 61 TCACCAAGTACGCGCGATAGCTTTTGGCGCTGAGACAGGGAACAATTACAGTTCAAGTACT 120

Qy 121 CAAAAAGCGCAACCTATAAAGCATATAAAGTTTTTGTATGACAGAAATAGATATGCAAT 180

121 DB CAAAAGGCGCAACCTATAAAGCATATAAAGTTTGTATGAGAAATAGATAATGCAAT 180
181 QY GTATCTGATTCGAATAAAGAGTGGAGCTCTCTATTAATTTCTCAAGGTAAAGAGCTGAG 240
181 DB GTATCTGATTCGAATAAAGAGTGGAGCTCTCTATTAATTTCTCAAGGTAAAGAGCTGAG 240
241 QY TATAAGCTTCAACTGATTTTAAATTTCTCTTTTACGACAACTACTAATGAGGAGGAACA 300
241 DB TATAAGCTTCAACTGATTTTAAATTTCTCTTTTACGACAACTACTAATGAGGAGGAACA 300
301 QY TATGTAACATAAAGAGATCTGCGTCAGCAAAATGAGATTGCGACATGGCTTAATCTATA 360
301 DB TATGTAACATAAAGAGATCTGCGTCAGCAAAATGAGATTGCGACATGGCTTAATCTATA 360
361 QY TCAGCTAATCTACACAGCTTTTCCACTGTTACTGAGTCAAAATATGATGGTACTGAGGTT 420
361 DB TCAGCTAATCTACACAGCTTTTCCACTGTTACTGAGTCAAAATATGATGGTACTGAGGTT 420
421 QY ATTAATGTTTCCCAATATGGAATATTAATGTTTCTAGCACTGTTAATATGAGAGCTGTA 480
421 DB ATTAATGTTTCCCAATATGGAATATTAATGTTTCTAGCACTGTTAATATGAGAGCTGTA 480
481 QY ATTATGTTTACATCTGTAACCTCCAAATGCTACTATTCATGAAAGATATCTGATGCGACA 540
481 DB ATTATGTTTACATCTGTAACCTCCAAATGCTACTATTCATGAAAGATATCTGATGCGACA 540
541 QY TGGGAGATGCGTGGGAAAGAACTGTAGTCAAAAAAGCTACTCGGTTGGTGATACAGTC 600
541 DB TGGGAGATGCGTGGGAAAGAACTGTAGTCAAAAAAGCTACTCGGTTGGTGATACAGTC 600
601 QY AAATATATCTATCTTATAGAATGCAAGTCAATATCATGGTACAGAAAAAGTGATCAA 660
601 DB AAATATATCTATCTTATAGAATGCAAGTCAATATCATGGTACAGAAAAAGTGATCAA 660
661 QY TATGTTATAAGGATATCTGCAATCTGCTCTGAGTTCATTTGAACGAGGCTCTTAT 720
661 DB TATGTTATAAGGATATCTGCAATCTGCTCTGAGTTCATTTGAACGAGGCTCTTAT 720
721 QY GAAGTAATCTATCTGATGATGAGGAAATATTAACAATCTACTCAAGGTTCGGAAAAA 780
721 DB GAAGTAATCTATCTGATGATGAGGAAATATTAACAATCTACTCAAGGTTCGGAAAAA 780
781 QY GCACTGGGAGATATACCTGTTAGAGGAAATATTAATTTACGATTAATTCGGTGG 840
781 DB GCACTGGGAGTATACCTGTTAGAGGAAATATTAATTTACGATTAATTCGGTGG 840
841 QY GCAGCTACCAATCTCCAAACCGGAAATCTCAAAATGAGCTAATGATGATCTTTTAT 900
841 DB GCAGCTACCAATCTCCAAACCGGAAATCTCAAAATGAGCTAATGATGATCTTTTAT 900
901 QY AAGGAAATAAATAAATCAAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960
901 DB AAGGAAATAAATAAATCAAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960
961 QY GGTTCAGCTGATTTACAGAAATACAAATGCGGACCATCAACCCCAATCTAGCAAT 1020
961 DB GGTTCAGCTGATTTACAGAAATACAAATGCGGACCATCAACCCCAATCTAGCAAT 1020
1021 QY GATGACCCAGGTCAAAAGTAAACAGTGGGAGTGGTCAAAATTAATTAATAAATAATGAT 1080
1021 DB GATGACCCAGGTCAAAAGTAAACAGTGGGAGTGGTCAAAATTAATTAATAAATAATGAT 1080
1081 QY GGTTCACAAAAGCTTCAATTAAGGTGCTATATTTGTTTAAAGAAATGCTACGGTCAA 1140
1081 DB GGTTCACAAAAGCTTCAATTAAGGTGCTATATTTGTTTAAAGAAATGCTACGGTCAA 1140
1141 QY TTTCTAACTTTAAGCATCAAAATACGTTGAATGGGACAGAACTAATGCAACAGAA 1200
1141 DB TTTCTAACTTTAAGCATCAAAATACGTTGAATGGGACAGAACTAATGCAACAGAA 1200
1201 QY TATACACAGGACGATGGTATATTAATACCATTAACAGGCTTGAAGAGGATACATACTAT 1260

1201 DB TATACACAGGACGATGGTATATTAATTAACCATTAACAGGCTTGAAGAGGATACATACTAT 1260
1261 QY CTAGTTGAGAAAAAGGCTCCCTTAGGTACAATTTGTTAGATACTCTCAGAAAGGTTATT 1320
1261 DB CTAGTTGAGAAAAAGGCTCCCTTAGGTACAATTTGTTAGATACTCTCAGAAAGGTTATT 1320
1321 QY TTAGGAGATGAGGCACTGATACGACTAAATTCAGATAACCTTTTAGTTAACCCAACTGTT 1380
1321 DB TTAGGAGATGAGGCACTGATACGACTAAATTCAGATAACCTTTTAGTTAACCCAACTGTT 1380
1381 QY GAAAAATAACAAAGTACTGAGTTCCCTTCAACAGGTTGTTATGTTACAAATTTTCTAC 1440
1381 DB GAAAAATAACAAAGTACTGAGTTCCCTTCAACAGGTTGTTATGTTACAAATTTTCTAC 1440
1441 QY ATTATAGGTGCAATTTTAGTAATAGGACGAGTATCGTCTGTTGCTCGTCTGTTTAA 1500
1441 DB ATTATAGGTGCAATTTTAGTAATAGGACGAGTATCGTCTGTTGCTCGTCTGTTTAA 1500
1501 QY CGTTCTTAA 1509
1501 DB CGTTCTTAA 1509

RESULT 2
US-10-206-576-217
; Sequence 217, Application US/10206576
; Publication No. US20030017495A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/206,576
; FILING DATE: 29-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/071,035
; FILING DATE: 1998-05-04
; APPLICATION NUMBER: US 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: US 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: US 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB369PID1
; INFORMATION FOR SEQ ID NO: 217:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1983 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 217:
US-10-206-576-217

Query Match 4.3%; Score 64.8; DB 13; Length 1983;
Best Local Similarity 52.1%; Pred. No. 0.0013;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAGCTAATGCACAGATATACACAGGAGCAGATGGTATATATACCATTCAGGC 1239
Db 1660 ACAAAGCTGAAGCAACTACTTTTACAAACCGCTGATGGATTAGTTGATATACAGGG 1719
QY 1240 TTGAAAGAGGTACATCTATCTAGTTGAGAAAAGGCTCCCTTAGGTTACAATTTGTTA 1299
Db 1720 CTTAATACGGTACCTATTATTATTAGAGAACTGTAGTCTCTGATGATTGTTGTTA 1779
QY 1300 GATACTCTCAGAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 1780 ACAATCGGATTGAATTTGGTCAATGAACAATCATATATGCGCAA-----CAGAAAC 1833
QY 1360 CTTTGTAGTTAACCACTGTTGAATAACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419
Db 1834 CTAGTTTCCAGAAAAGTACCAACAAACAAAGGTACTTACCTTCAACAGGTGGC 1893
QY 1420 ATTGGTACAAATTTTCTACATTATAGTGTCAATTTTAGTAATAGGAGCAGGTATCGTG 1479
Db 1894 AAAGGAATCTACGTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTCAGGAGTCTAC 1953
QY 1480 CTTGTTGCTCGTCTGCTGTTACGTTCTTAA 1509
Db 1954 TTTGCTAGACGTAGAAAAGAAATGCTTAA 1983

RESULT 3

US-10-333-002-28
; Sequence 28, Application US/10333002
; Publication No. US20040071729A1
; GENERAL INFORMATION:
; APPLICANT: Adderson, Elisabeth
; APPLICANT: Bohnsack, John
; TITLE OF INVENTION: Group B Streptococcus Polypeptides Nucleic Acids and Therapeutic
; FILE REFERENCE: 2511-1-001 (SJ-0039)
; CURRENT APPLICATION NUMBER: US/10/333,002
; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: PCT/US01/24795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/634,341
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 2199
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-10-333-002-28

Query Match 4.3%; Score 64.8; DB 12; Length 2199;
Best Local Similarity 52.1%; Pred. No. 0.0013;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;
QY 1180 ACAGAGCTAATGCACAGATATACACAGGAGCAGATGGTATATATACCATTCAGGC 1239
Db 275 ACAAAGCTGAAGCAACTACTTTTACAAACCGCTGATGGATTAGTTGATATACAGGG 334
QY 1240 TTGAAAGAGGTACATCTATCTAGTTGAGAAAAGGCTCCCTTAGGTTACAATTTGTTA 1299
Db 335 CTTAATACGGTACCTATTATTATTAGAGAACTGTAGTCTCTGATGATTGTTGTTA 394
QY 1300 GATACTCTCAGAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 395 ACAATCGGATTGAATTTGGTCAATGAACAATCATATGCGCAA-----CAGAAAC 448
QY 1360 CTTTGTAGTTAACCACTGTTGAATAACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419
Db 449 CTAGTTTCCAGAAAAGTACCAACAAACAAAGGTACTTACCTTCAACAGGTGGC 508
QY 1420 ATTGGTACAAATTTTCTACATTATAGTGTCAATTTTAGTAATAGGAGCAGGTATCGTG 1479
Db 509 AAAGGAATCTACGTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTGAGGAGTCTAC 568

QY 1480 CTTGTTGCTCGTCTGCTGTTTACGTTCTTAA 1509
Db 569 TTTGCTAGACGTAGAAAAGAAATGCTTAA 598

RESULT 4

US-09-070-927A-45
; Sequence 45, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15614 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-070-927A-45

Query Match 4.3%; Score 64.8; DB 9; Length 15614;
Best Local Similarity 52.1%; Pred. No. 0.0035;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;
QY 1180 ACAGAGCTAATGCACAGATATACACAGGAGCAGATGGTATATATACCATTCAGGC 1239
Db 13724 ACAAAGCTGAAGCAACTACTTTTACAAACCGCTGATGGATTAGTTGATATACAGGG 13783
QY 1240 TTGAAAGAGGTACATCTATCTAGTTGAGAAAAGGCTCCCTTAGGTTACAATTTGTTA 1299
Db 13784 CTTAATACGGTACCTATTATTATTAGAAAAGTCTGATGCTCTCTGATGATTGTTGTTA 13843
QY 1300 GATACTCTCAGAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 13844 ACAATCGGATTGAATTTGGTCAATGAACAATCATATGCGCAA-----CAGAAAC 13897
QY 1360 CTTTGTAGTTAACCACTGTTGAATAACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419
Db 13898 CTAGTTTCCAGAAAAGTACCAACAAACAAAGGTACTTACCTTCAACAGGTGGC 13957

Qy 1420 ATTGGTACACAATTTTCTACATATAGTGTCAATTTTGTATATAGGAGCAGGTATCGTG 1479

Db 13958 AAGGAATCTACGTTTACTTTAGGAAGTGGCGAGCTTGTCTACTATTGCAGAGCTTAC 14017

Qy 1480 CTTGTTGCTGCTGCTGCTTTACGTTCTTAA 1509

Db 14018 TTTGCTAGACGTAGAAAAGAAATGCTTAA 14047

RESULT 5

```

US-10-661-809-12
; Sequence 12, Application US/10661809
; Publication No. US2004010191A1
; GENERAL INFORMATION:
; APPLICANT: HOOK, Magnus
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FROM
; FILE REFERENCES: F07741US01/BAS
; CURRENT APPLICATION NUMBER: US/10/661,809
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: 60/410303
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1881
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-661-809-12

```

Query Match 4.2%; Score 63.2; DB 17; Length 1881;
Best Local Similarity 52.5%; Pred. No. 0.0027;
Matches 165; Conservative 0; Mismatches 143; Indels 6; Gaps 1;

1180	QY	ACAGAAGCTAATGCAACAGAAATATACAACAGGACAGATGGTATTAATTAACCAATTACAGGC	1239
1561	Db	ACAAAGCTGAAGCAACTACTTTTACAACAACGGCTGATGGATTAGTTTGATATCACAGG	1620
1240	QY	TTGAAAGAGGTACATACTACTCTAGTTTGAGAAAAGGCTCCCTTAGGTTACAATTGTTTA	1299
1621	Db	CTTAATACGGTACCTATTATTAGAAGAACTGTAGCTCCTGATGATTGTCTTGTTA	1680
1300	QY	GATAACTCTCAGAAGGTTATTTTAGGAGATGGAGCCACTGATPACGACTAATTCAGATAAC	1359
1681	Db	ACAAATCGGATGGAATTTGGTCAATGAACAATCATATGGCACAA-----CAGAAAC	1734
1360	QY	CTTTTAGTTAAACCAACTGTTTGAATAACAAGGTACTAGCTGCCTTCAACAGGTGGT	1419
1735	Db	CTAGTTTCACGAGAAAAGTACCACAACAACAAAGGTACCTTACCTTCAACAGGTGGC	1794
1420	QY	ATTGGTACACAATTTTCTACATTATAGGTGCAATTTTAGTAATAGGACAGGTATCGTG	1479
1795	Db	AAGGAATCTACGTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTTCAGGAGTCTAC	1854
1480	QY	CTTGTGCTCGTCG	1493
1855	Db	TTTGCTAGCGTAG	1868

RESULT. T. 6

US-10-282-122A-15988
; Sequence 15988, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John

```

/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITFA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 15986
/ LENGTH: 1137
/ TYPE: DNA
/ ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15986

```

Query Match 3.9%; Score 59.2; DB 13; Length 1137;
Best Local Similarity 43.7%; Pred. No. 0.015;
Matches 262; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

QY	90	TGAGACAGGACAATTTACAGTTC	149
DB	30	TGGGCAAGGCCAGAAAGCTATAAAGATGGCTCTCTTGGTTAAAAAACTGAAAGAAAATAA	89
QY	150	AGTTTTTCATGCAGAAATAGATAATGCAATGTATCTGATTCGAAATAAGATGGAGCTTC	209
DB	90	AAATTTTCAAGCAAAAGTATGTGTTACAGCACAGCAGGGAGATGCTAGATCAAGTTCT	149
QY	210	TTATTTAAATTCCTCAAGGTAAGAGCTTGAGTATAAAGCTTCAACTGATTTTAAATTCCT	269
DB	150	TAGTTTGTTCAAATATAGAGCCAGACTTTTGATTTTCAATATAAATGAAATAAGCAATCTCT	209
QY	270	TTTTACGACAACTACTAATGAGGGAGACAATATGTAACTAAAAAAAGATCTGCGTCAGC	329
DB	210	TACACAGGATACTTCAAGTGTGTTTACATGGGTTTAAAGCGAAATATTTTCTGCTGAAAGACC	269
QY	330	AAATCAGATTCGCATCGGGCTMAATCTATATCAGCTAATACCTACACACAGTTTCCACTGT	389
DB	270	AGATATTATTTAGTACATGGAGATACAACTACAAAGTTTGGAGCATCACTTCGAGCCTT	329
QY	390	TACTGAGTCAAAATGATGGTACTGAGGTTATAATGTTTCCCAATATGGATATTATTA	449
DB	330	TTATGAAAAGATAGCTATTGGGCATGTTGAGGCTGGACTTTAGAACTTATGATAAATAATTT	389
QY	450	TGTTTTCTAGCACCTGTTAATAATCGAGCTGTAATTTATGTTTACATCTGTAACTCCAAATGC	509
DB	390	TCCTTTTCTGAAAGAGTAATAATAGAAAATTACTCGAGCAATTCGATGATATGCATTTTGC	449
QY	510	TACTATTTCATGAAAAGAAATCTATGTCGACATGGGAGATGGTGGTGGAAAAAAGCTGTAGA	569
DB	450	ACCGACAGTAAATCTTAAAAATAATCTTTTAAAGAGAAGGTGTAAAAAGAAAAAATAATTT	509

[illegible]

RESULT 7

```

US-10-094-240-10/c
; Sequence 10, Application US/10094240
; Publication No. US20030082637A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
; FILE REFERENCE: N8289
; CURRENT APPLICATION NUMBER: US/10/094,240
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 10/056,405
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-094-240-10

```

RESULT 8

```

US-10-056-405-10/c
; Sequence 10, Application US/10056405
; Publication No. US20030166013A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: MOSQUITO OLFATORY GENES, POLYPEPTIDES, AND METHODS OF
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: N7841
; CURRENT APPLICATION NUMBER: US/10/056,405
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-056-405-10

```

RESULT 9

```

US-10-349-680-149/c
; Sequence 149, Application US/10349680
; Publication No. US20030176654A1
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail
; APPLICANT: Chen, Ellison
; APPLICANT: Glass, Jennifer
; APPLICANT: Glass, John
; APPLICANT: Heiner, Cheryl
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA

```

```

; TITLE OF INVENTION: UREALYTICUM
; FILE REFERENCE: UAB-13403/22
; CURRENT APPLICATION NUMBER: US/10/349,680
; CURRENT FILING DATE: 2003-01-23
; PRIOR APPLICATION NUMBER: US 09/601,198
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: PCT/US99/01972
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: US 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 149
; LENGTH: 14066
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-10-349-680-149

Query Match      3.3%; Score 50.4; DB 15; Length 14066;
Best Local Similarity 42.8%; Pred. No. 4.1;
Matches 313; Conservative 0; Mismatches 416; Indels 3; Gaps 1;

QY 435 ATATGGATATTATTATGTTTCTAGCAGCTGTTTAATAATGGAGCTGTAAATTATGTTTACATC 494
DB 5284 AAATGTTAGTTCTAATGATTCTATAAATACTACAACAACAACAGCTTAATATTGATTTGA 5225
QY 495 TGTAACTCCAAATGCTACTATTTCATGAAGAAGAACTGATCGGCAGATGGGGAGATGTTGG 554
DB 5224 TGGAAATTCACGCACTTGAAATTAATAAAAAAATTCAAATGGTATATATCTTCAAAATGACGG 5165
QY 555 TGGAAAACTGTAGATCAAAAACGCTACTCGTTTGGTGATACAGTCAAAATATACTATTATC 614
DB 5164 TGAAGAAATTTTAAAGTGATCAAAAAACATTCGTTTGGGCTAATAAACACATATAGTTTCGA 5105
QY 615 TTATAAGATGTCAGTCAATTATTCATGTGCACAGAAAAGTGATCAATATGTTATAAAGGA 674
DB 5104 ATTGCTCAATTTTAAACCAATAATCGTAATAATACACTAAAAAGAGTAAGAAATTTAATGA 5045
QY 675 TACTATGCCATCTGCTTCTGTAGTTGATTTGAACGAGAGGTCATTATCAAGTCAACTATTAC 734
DB 5044 TGATAATAAACATCAATAATTTTCCATTTTAAAAATTGAATAGCTAGCTGATTTATGTT 4985
QY 735 TGATGGATCAGGGAAATATTCAACTCTPACTCAAGGTCGGAAGAAACAACTCGGGAAGTA 794
DB 4984 TAATAAAACAACAATCGATTAGTATAAGTTCAATAATTGAGCCAAATTCGAAGAGCGAAAA 4925
QY 795 TAACCTGTTAGAGGAAATAATAATTTTCAGATTACTATTTCGTTGGGCAGCTACCAATAC 854
DB 4924 TAATTTACAATCAACACAAATTAGATTATTAATAATGATCCTCGATTAATGTTTAAAGTAA 4865
QY 855 TCCAACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTTTTTATAAGGGAATAAATAC 914
DB 4864 TGAAGAAGAGCAATCATTATTTATGTTGAGAAAT--TTATCTGTAAGACCAAAAGTTAA 4808
QY 915 AATTCAGTCACTTATACAGAGTATTAAAGAGTGGAGCTAAACCGAGTTTCAGCTGATTT 974
DB 4807 AATTTGAGGAGTCAAAAATAATTTTAGAAACAACATTCAAATACTAGCAATTAACCAAGA 4748
QY 975 ACCAGAAAAATACAAAACATTCGCGACCACTCAACCCCAATCTAGCAATGATGACCCAGTCA 1034
DB 4747 TACTATTATTAAACAGTATTACTTTTATTATAAACCAATAAAGCTGCTACGAATATTGG 4688
QY 1035 AAAAGTAACTGAGGGATGGTCAAAATTACTATAAAAAAAATGATGGTTTCCAAAAAGC 1094
DB 4687 AATAGATAATAGCAATAAAAAATATAACATAACAAACAATAAATCCATTAATAAATTGA 4628
QY 1095 TTCAATCAAGGTGCTATATTTGTTTAAAGATGCTACGGGTCAATTTCTAAACTTTAA 1154
DB 4627 TAATAATTTTAGTGTCAATGGTCTCTGTTCAAGTGATACTAATAAACGTAATTACTTTGGA 4568
QY 1155 CGATACAAAATAA 1166
DB 4567 ATATAATAATAA 4556

```


[illegible]

```

RESULT 11
US-10-087-464-42
; Sequence 42, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishti, Athar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: SI237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 42
; LENGTH: 3996
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-10-087-464-42

```

	Query Match	3.3%	Score 50;	DB 15;	Length 3996;
	Best Local Similarity	47.5%;	Pred. No. 2.7;		
	Matches 149;	Conservative 0;	Mismatches 165;	Indels 0;	Gaps 0;
QY	366	TAATACTACACCAAGTTTCCACTGTTACTGAGTCAATAATGATGGTACTCAGGTTATTAA	425		
Db	1881	TAATAATAATGATGAGAAATTTGTGATAATAATAAACCATAAATAATAATTAATATAATAA	1940		
QY	426	TGTTTCCCAATATGGATATTATTATGTTTCTAGCACGTGTTAAATAAGGAGCTGTAAATTAT	485		
Db	1941	TAATAATAATTATGGTAAACAATAATAATAACAACAATAATAATAAGGACAATAATAATAA	2000		
QY	486	GGTTACATCTGTAACTCCAAATGCTACTATTCTATGAAAGAATACTGATCGCACATGGGG	545		

[illegible]

RESULT 12
 US-10-206-576-219
 ; Sequence 219, Application US/10206576
 ; Publication No. US20030017495A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et al.
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 ; NUMBER OF SEQUENCES: 497
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-R
 ; COMPUTER: Dell Latitude
 ; OPERATING SYSTEM: Windows 98
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/206,576
 ; FILING DATE: 29-Jul-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/071,035
 ; FILING DATE: 1998-05-04
 ; APPLICATION NUMBER: US 60/046,655
 ; FILING DATE: 1997-05-16
 ; APPLICATION NUMBER: US 60/044,031
 ; FILING DATE: 1997-05-06
 ; APPLICATION NUMBER: US 60/066,009
 ; FILING DATE: 1997-11-14
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hyman, Mark J.
 ; REGISTRATION NUMBER: 46,789
 ; REFERENCE/DOCKET NUMBER: PB369P1d1
 ; INFORMATION FOR SEQ ID NO: 219:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1687 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 219:
 US-10-206-576-219

	Query Match	3.3%	Score 49.8	DB 13	Length 1687
	Best Local Similarity	62.4%	Pred. No. 1.9		
	Matches	79	Conservative	0	Mismatches 47; Indels 0; Gaps 0;
Qy	1180	ACGAGAGCTTAATCGACACAGAAATATACAA	CGGAGCAGATGGTATAATTACCATTCACGGC	1239	
Db	1466	ACAAAAGCTGAAGCAACTACTTTT	TACAACAAACGGCTGATGGATTAGTTGATATACAGGG	1525	
Qy	1240	TTGAAAGAGGTCATACATCTACTAGTTG	AAGAAAAAGCTCCCTTAGGTTACAATTTGTTA	1399	
Db	1526	CTTAAATACGCGTACCTATTATTTT	AAGAAAGAACTGTAGCTCCGATGATTATGCTTTGTTA	1585	

```
QY 1300 GATAA 1304
Db 1586 ACAA 1590

RESULT 13
US-10-282-122A-35506
; Sequence 35506, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35506
; LENGTH: 4997
; TYPE: DNA
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-35506

Query Match 3.2%; Score 48.6; DB 13; Length 4997;
Best Local Similarity 43.5%; Pred. No. 6;
Matches 271; Conservative 0; Mismatches 349; Indels 3; Gaps 1;

QY 28 TTAGTGGCGAGTTAGCATTTGGTATGGCTGATCACCAGTACCGCGATAGCTTTGCC 87
Db 1741 TCAGATCAAAACNAGCAATGATACAGCAGTCTTAAACCAAAATCATTAGTCAATAAT 1800

QY 88 GCTGAGACAGGCAATTTACAGTTTCAAGATACCTCAAAAAGCGCAACCTATAAGCATAT 147
Db 1801 AACAAATGAAGCCACTCAAGAGAGAGAGATGAGCTTTAGCCAAATTTGATGAGCAGCA 1860

QY 148 AAGTTTTTGAATGAGAAATAGATAATGAAATGATCTGATTCGAAATAAAGATGAGCT 207
Db 1861 AAACAAGCTAAAGCTGCGATAGATGCTGCAACTACAAATAATGC---TGTAGATGAAGCA 1917

QY 208 TCTTATTTAATCTCTCAAGGTAAGAGCTGAGTATATAAGCTTCACTGATTTTAATCT 267

QY 1918 ACAAACAATAACTACTACAATTTATTTCTGGAATACTTCTCTGATACGGTGAAGAAAGCAGCA 1977
Db 268 CTTTTTACGCAACTACTAATGAGGAGGAGAACATATGTACTTAATAAAGATAGTGGCTCA 327
Db 1978 GCAAGAAAGCGATTGATGATGAGCAGCAACTGCTAAAAAAGAGCGATAAATAATACGTCA 2037

QY 328 GCAAAATGAGATTGCGCATGGCTAAATCTATATCAGCTAATCTACTACACAGTTTCCACT 387
Db 2038 GATGCAACACAGAAGAAAGATGAGCAATAGCGAAAGTTGATGCGAGCTGTAACAGCT 2097

QY 388 GTTACTGAGTCAAAATAATGATGCTAGCTGAGTTTATTAATGTTTCCCAATATGGATATTAT 447
Db 2098 GCAAAACAGCAATTTACACAAGCAACAACAATAATGATAATGTAGACCAACAATAATAGC 2157

QY 448 TATGTTTCTAGCACTGTTAATAATGAGCTGTAATTAATGTTTCTTACATCTGTAATCCCAAT 507
Db 2158 GGTACTTCGACTATTACTTGTATACCAAGAGTTTACTAAAAAAGCAGCAGCAAGAAA 2217

QY 508 GCTACTATTTCATGAAAAGATACTGATGCCACATGGGAGATGGTGGTGAAGAACTGTGA 567
Db 2218 GCAATTGATGATGAGTCGTTGCTAAGAAAGCAGCAATTGATCTGTTGCCGATGTACA 2277

QY 568 GATCAAAAACGCTACTCGGTTGGTGATACAGTCAAAATATATCTATTACTTTATAAGATGCA 627
Db 2278 GATGAAGAAAAACAGGCAGCTAAAGATAAAGTTGATGCTGAAGCTACCAAGCAAGCA 2337

QY 628 GTCAATTATCATGTCGTACAGAAA 650
Db 2338 GCGATTGATCAAGCTACTACAAA 2360

RESULT 14
US-09-137-531-13
; Sequence 13, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-137-531-13

Query Match 3.2%; Score 48.4; DB 9; Length 3666;
Best Local Similarity 44.5%; Pred. No. 5.6;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 302 ATGTAACTAAAAAGTACTCGCTCAGCAAAATGAGATTCGGACATGGGCTAATCTATAT 361
Db 2768 ATGTAACTGCAACAACTGTTGATCTGCAACTGTTTCATTAAAGATAGTGCATAAT 2827

QY 362 CAGCTAATACTACACAGTTTCCACTGTTTACTGAGTCAAAATATGATGCTACTGAGGTTA 421
Db 2828 CATTATCTTACATTAGTTTGAACCTGGTGCTTAATACAGGTGTTTGTCTACACCTGTT 2887
```

```
QY 422 TTAATGTTTCCCAATATGGATATATATATGTTTCTAGCACTGTTAATAATGGAGCTGTA 481
Db 2888 AAGCTGTGTACATTATCTTAACTGCTGGTACATTAAACAGTTACTTATGACAGATGCTA 2947
QY 482 TTATGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAAAGAAATACTGATGCCAGAT 541
Db 2948 ABAATGCTGAGGTGTTGCTGAAAAATATTACTGCTAGCGTAAACNTTAAAGAAAACTACTG 3007
QY 542 GGGAGATGGTGTGGAAGAACTGTAGATCAAAAAAGTACTCGGTTTGGTGATACAGTCA 601
Db 3008 GAGCAATTACTTCTGATACATTTACACAAAGGTGTATTACCATCAGCAGCTTACAGCAGCTG 3067
QY 602 AATATATCTATTACTTATAAGAAATGCAAGTCAATTTATCATGTTGATGAGTAAAGTGTATCAAT 661
Db 3068 AATATATCTTCTTAAATCAATTTGCTGAGATTATACATTTGCAACAGGTGAAGGATTCACCT 3127
QY 662 ATGTTTAAAGGATACATGCCATCTGCTTCTGCTAGTTGATTTGCAACGAAAGGCTCTTATG 721
Db 3128 TAAATATTGATAATGCTGGTGCTCAAGTAATTAACCTAGCAGGTAAAAAAGGTGCACAAG 3187
QY 722 AAGTAATCTATTACT 735
Db 3188 GTGTAGCTGATGCT 3201
```

RESULT 15

```
US-09-137-531-14
; Sequence 14, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3666
US-09-137-531-14
```

```
Query Match 3.2%; Score 48.4; DB 9; Length 3666;
Best Local Similarity 44.5%; Pred. No. 5.6;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 302 ATGTAATAAAAAAGATACCTGCGTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATAT 361
Db 2768 ATGTAAGTGCACAACTGTTGATACATGCAACTGTTTCATTAAAAAGATAGTGCAAAATAAT 2827
QY 362 CAGCTAATACACACAGTTTCACCTGTTACTGAGTCAAAATATGATGGTACTGAGGTTA 421
Db 2828 CATTATCTCTTACATTATTGAAACTGGTGTCTAATACAGGTGATTTTGTACAACTGTTTC 2887
QY 422 TTAATGTTTCCCAATATGGATATATATATGTTTCTAGCACTGTTAATAATGGAGCTGTA 481
Db 2888 AAGCTGTGTACATTATCTTAACTGCTGGTACATTAACAGTTACTTATGACAGATGCTA 2947
```

```
QY 482 TTATGTTTACATCTGTAACTCCAAATGCTACTATTTCATGAAAGAAATACTGATGCCAGAT 541
Db 2948 ABAATGCTGAGGTGTTGCTGAAAAATATTACTGCTAGCGTAAACNTTAAAGAAAACTACTG 3007
QY 542 GGGAGATGGTGTGGAAGAACTGTAGATCAAAAAAGTACTCGGTTTGGTGATACAGTCA 601
Db 3008 GAGCAATTACTTCTGATACATTTACACAAAGGTGTATTACCATCAGCAGCTTACAGCAGCTG 3067
QY 602 AATATATCTATTACTTATAAGAAATGCAAGTCAATTTATCATGTTGATGAGTAAAGTGTATCAAT 661
Db 3068 AATATATCTTCTTAAATCAATTTGCTGAGATTATACATTTGCAACAGGTGAAGGATTCACCT 3127
QY 662 ATGTTTAAAGGATACATGCCATCTGCTTCTGCTAGTTGATTTGCAACGAAAGGCTCTTATG 721
Db 3128 TAAATATTGATAATGCTGGTGCTCAAGTAATTAACCTAGCAGGTAAAAAAGGTGCACAAG 3187
QY 722 AAGTAATCTATTACT 735
Db 3188 GTGTAGCTGATGCT 3201
```

Search completed: September 13, 2004, 08:40:15
Job time : 733.222 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2004, 22:26:26 ; Search time 604.354 Seconds
(without alignments)
10607.240 Million cell updates/sec

Title: US-10-009-254-1
Perfect score: 1509
Sequence: 1 atgaaaagaagaatgattca.....gtcgtggttaacgtttcttaa 1509

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1509	100.0	1509	5	AAD02390
2	1509	100.0	1509	6	ABK11585
3	64.8	4.3	1983	2	AAX20108
4	64.8	4.3	1983	6	ABN98093
5	64.8	4.3	1983	7	ACA88057
6	64.8	4.3	1983	7	ABX61663
7	64.8	4.3	2199	6	ABK11591
8	64.8	4.3	15614	2	AAX12982
9	64.8	4.3	15614	6	ABS98777
10	59.6	3.9	11922	3	AAA70187
11	59.2	3.9	1137	7	ACA28118
12	56.6	3.8	969	6	ABQ39490
13	56.6	3.8	969	6	ABQ39491
14	55.4	3.7	2408	4	ABL28164
15	55.2	3.7	32392	6	ABL56203
16	54	3.6	408	4	ABL28165
17	53.6	3.6	2535	4	ABL25476
18	53	3.5	4985	6	ABQ75107
19	53	3.5	4985	9	ACF79720
20	53	3.5	6033	3	AAA70152
21	51.2	3.4	2703	6	ABN67916
22	51.2	3.4	110000	6	ABN71527_13
23	50.4	3.3	14066	2	AAX99556

24	50.4	3.3	14067	7	ACA52811	ACA52811	Prokaryot
25	50.2	3.3	7143	3	AAA70250	AAA70250	Plasmodi
26	50	3.3	3996	6	AAD47004	AAD47004	Plasmodi
27	49.8	3.3	1687	2	AAX20109	AAX20109	Enterococ
28	49.8	3.3	1687	6	ABN98094	ABN98094	E faecali
29	49.8	3.3	1687	7	ACA88058	ACA88058	E faecal
30	49.8	3.3	1687	7	ABX61664	ABX61664	Enterococ
31	49.6	3.3	3837	3	AAA70211	AAA70211	Plasmodi
32	48.6	3.2	4997	7	ACA47636	ACA47636	Prokaryot
33	48.6	3.2	48551	6	AAAS20800	AAAS20800	Clostridi
34	48.4	3.2	4197	2	AAQ99430	AAQ99430	B. sphaer
35	47.8	3.2	740	4	AAC85918	AAC85918	rCP41 CDN
36	47.8	3.2	1083	5	AAAS76745	AAAS76745	DNA encod
37	47.8	3.2	8244	5	AAAS6529	AAAS6529	DNA encod
38	47.6	3.2	11143	4	ABL12834	ABL12834	Drosophil
39	47.4	3.1	1037	3	AAA59242	AAA59242	Exons E,
40	47.4	3.1	1472	3	AAA59241	AAA59241	Exons D,
41	47.4	3.1	2428	7	ADA89818	ADA89818	Staphyloc
42	47.4	3.1	3945	2	AAX77593	AAX77593	S. aureus
43	47.4	3.1	4047	4	AAAS2261	AAAS2261	Staphyloc
44	47.4	3.1	4047	7	ACF73996	ACF73996	Staphyloc
45	47.4	3.1	4050	4	AAAS5402	AAAS5402	Staphyloc

ALIGNMENTS

RESULT 1
AAD02390
ID AAD02390 standard; DNA; 1509 BP.
XX
AC AAD02390;
XX
DT 24-APR-2001 (first entry)
XX
DE Virulent group B Streptococcus agalactiae spbl DNA.

XX
Type III virulent group B; spbl; cell wall bound protein; antibacterial;
KW Immunisation; group B streptococci; GBS infection; vaccine; bacteraemia;
KW pneumonia; meningitis; endocarditis; osteoarticular infection; ds.

OS Streptococcus agalactiae.
FH Key Location/Qualifiers
FT CDS 1..1509
FT FT /tag= a
FT FT /product= "S. agalactiae spbl protein"
FT sig_peptide 19..87
FT FT /tag= b
FT mat_peptide 88..1506
FT FT /tag= c
FT FT /product= "S. agalactiae mature spbl protein"

WO200078787-A1.

28-DEC-2000.

21-JUN-2000; 2000WO-US017082.

21-JUN-1999; 99US-0140084P.

(UTAH) UNIV UTAH RES FOUND.

Adderson E, Bohnsack J;

WPI; 2001-102693/11.

P-PSDB; AAY72357.

Polynucleotide from spbl and 2 genes derived from virulent Group B streptococci, polypeptide encoded by the polynucleotide useful as vaccine for immunizing a mammal against the streptococcal infection.

Claim 9; Page 20-23; 34pp; English.

XX The present sequence is sbpl gene from type III virulent group B
 CC Streptococcus agalactiae. The sbpl protein has the characteristics of a
 CC cell wall bound protein and has antibacterial activity. The N-terminus
 CC of the sbpl protein is a hydrophilic, basic stretch of 6 amino acids
 CC followed by a 23 amino acid hydrophobic, proline rich core, consistent
 CC with a signal peptide. The hydrophilic mature protein terminates in
 CC atypical LexTG domain that immediately precedes a hydro- phobic 20 amino
 CC acid core and a short, basic hydrophilic terminus. The sbpl protein is
 CC used as a vaccine to immunise mammals against group B Streptococci (GBS)
 CC infection (e.g. bacteraemia, pneumonia, meningitis, endocarditis and
 CC osteoarticular infections). Determination of the gene products specific
 CC to type III-3 GBS is useful for diagnosing mammals infected or colonised
 CC by virulent GBS

XX Sequence 1509 BP; 534 A; 241 C; 298 G; 436 T; 0 U; 0 Other;

Query Match	100.0%;	Score 1509;	DB 5;	Length 1509;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1509;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 ATGAAAAAGAAATGATTCATTCGCTGTAGTGGCGAGTTTAGCAATTTGGTATGGCTGTA 60
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 61 TCACCAAGTACGCCGATAGCTTTTCCGCTGAGACAGGACAATACAGTTCAAGATACT 120
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 61 TCACCAAGTACGCCGATAGCTTTTCCGCTGAGACAGGACAATACAGTTCAAGATACT 120
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 121 CAAAAAGCGCACTATAAAGCATATAAAGTTTGTATGATGAGAAATAGATAATGCAAT 180
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 121 CAAAAAGCGCACTATAAAGCATATAAAGTTTGTATGATGAGAAATAGATAATGCAAT 180
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 181 GTATCTGATTCGAATAAAGATGGAGCTTCTTATTTAATTCCTCAAGSTAAAGAGCTGAG 240
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 241 TATAAGCTTCAACGTGATTTAATTCCTTTTACGACAATCTACTAATGAGGAGGAACA 300
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 241 TATAAGCTTCAACGTGATTTAATTCCTTTTACGACAATCTACTAATGAGGAGGAACA 300
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 301 TATGTAACTAAAAAGATACCTCGCTCAGCAATAGAGATTGGACATGGCTTAATCTATA 360
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 301 TATGTAACTAAAAAGATACCTCGCTCAGCAATAGAGATTGGACATGGCTTAATCTATA 360
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 361 TCAGTAATACACAGATTTTCACTGTTACTGAGTCAATATATGTTGTTACTGAGGTT 420
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 361 TCAGTAATACACAGATTTTCACTGTTACTGAGTCAATATATGTTGTTACTGAGGTT 420
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 421 ATTAATGTTTCCCAATATGGATATTTATGTTTCTAGCACTGTTAATATGAGCTGTA 480
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 421 ATTAATGTTTCCCAATATGGATATTTATGTTTCTAGCACTGTTAATATGAGCTGTA 480
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 481 ATTATGTTACATCTGTAACCTCAAATGCTACTATTTCATGAAAAGATACCTGATGCGACA 540
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 481 ATTATGTTACATCTGTAACCTCAAATGCTACTATTTCATGAAAAGATACCTGATGCGACA 540
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 541 TGGGAGATGGTGGTGAAGAACTGTAGATCAAAAACGTACTCGGTTGGTGTATACAGTC 600
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 541 TGGGAGATGGTGGTGAAGAACTGTAGATCAAAAACGTACTCGGTTGGTGTATACAGTC 600
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 601 AAATATACTATTACTTATAGAAATGACGTCAATTTATCATGTGTACAGAAAAGTATCA 660
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 601 AAATATACTATTACTTATAGAAATGACGTCAATTTATCATGTGTACAGAAAAGTATCA 660
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 661 TATGTTTAAAGGATACATGCAATCTGCTTCTGTAGTTGATTTGAACGAAGGCTTAT 720
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 661 TATGTTTAAAGGATACATGCAATCTGCTTCTGTAGTTGATTTGAACGAAGGCTTAT 720
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 721 GAAGTAACTATTACTGATGATCAGGGAATATTACAACTCACTCACTCACTCACTCACTCACT 780
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 721 GAAGTAACTATTACTGATGATCAGGGAATATTACAACTCACTCACTCACTCACTCACTCACT 780
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY	781	GCAACTGGGAAGTATAACCTGTTAGAGGAAAATAATAATTTACGATTACTATTCCGTTG	840
Db	781	GCAACTGGGAAGTATAACCTGTTAGAGGAAAATAATAATTTACGATTACTATTCCGTTG	840
QY	841	GCAGTCCCAATATCTCCAAACCGGAAATACCTCAAAATGGAGCTAATGATGACTTTTTTAT	900
Db	841	GCAGTCCCAATATCTCCAAACCGGAAATACCTCAAAATGGAGCTAATGATGACTTTTTTAT	900
QY	901	AAGGGAATAAATAAATCAATCAAGTCACTATACAGGAGTATTAAGAGTGGAGCTAAACCA	960
Db	901	AAGGGAATAAATAAATCAATCAAGTCACTATACAGGAGTATTAAGAGTGGAGCTAAACCA	960
QY	961	GGTTGAGCTGATTTACAGGAAAATACAAATTCGCGACCAATCAACCCCAATCTAGCAAT	1020
Db	961	GGTTGAGCTGATTTACAGGAAAATACAAATTCGCGACCAATCAACCCCAATCTAGCAAT	1020
QY	1021	GATGACCCAGCTCAAAAGTAAACAGTGGAGGATGGTCAAAATTAATAAAAAAATTTGAT	1080
Db	1021	GATGACCCAGCTCAAAAGTAAACAGTGGAGGATGGTCAAAATTAATAAAAAAATTTGAT	1080
QY	1081	GGTTCCCAAAAAGCTTCATTACAAGGTCTATATTTGTTTAAAGATGCTACGGGTCAA	1140
Db	1081	GGTTCCCAAAAAGCTTCATTACAAGGTCTATATTTGTTTAAAGATGCTACGGGTCAA	1140
QY	1141	TTTCTAAACTTTAACGATACAAATAACGTTGAATGGGGCACAGAGCTAATGCAACAGAA	1200
Db	1141	TTTCTAAACTTTAACGATACAAATAACGTTGAATGGGGCACAGAGCTAATGCAACAGAA	1200
QY	1201	TATACAACAGGACAGATGATATAATCCATTACAGGCTTGAAGAGGTACATCTAT	1260
Db	1201	TATACAACAGGACAGATGATATAATCCATTACAGGCTTGAAGAGGTACATCTAT	1260
QY	1261	CTAGTTTGAGAAAAGGCTCCCTTAGGTACAATTTCTGTAGATAAATCTCAGAGGTTTAT	1320
Db	1261	CTAGTTTGAGAAAAGGCTCCCTTAGGTACAATTTCTGTAGATAAATCTCAGAGGTTTAT	1320
QY	1321	TTAGGAGTGGAGCCACTGATACGACTAATTCAGATAAATCTTTAGTTAAACCACTGTT	1380
Db	1321	TTAGGAGTGGAGCCACTGATACGACTAATTCAGATAAATCTTTAGTTAAACCACTGTT	1380
QY	1381	GAAATAACAAGGTAAGTGGCTTCAACAGGTTGTTGGTACAAATTTTCTAC	1440
Db	1381	GAAATAACAAGGTAAGTGGCTTCAACAGGTTGTTGGTACAAATTTTCTAC	1440
QY	1441	ATTATAGTGCATTTTATAGTATAGGAGGAGTATCGTCTTGTGCTCGTCTCGTTTAA	1500
Db	1441	ATTATAGTGCATTTTATAGTATAGGAGGAGTATCGTCTTGTGCTCGTCTCGTTTAA	1500
QY	1501	CGTCTTAA	1509
Db	1501	CGTCTTAA	1509

RESULT 2

ABK11585
 ID ABK11585 standard; DNA; 1509 BP.

XX ABK11585;

XX AC

XX XX

XX 05-JUN-2002 (first entry)

XX S. agalactiae SbpI gene.

XX Extracellular matrix adhesion; Ems; ds; group B streptococcus; GBS;

XX DNA vaccine; SbpI; Sbp2; Rib; Lmb; Csa-ase; C protein alpha antigen;

XX neonatal bacterial infection; gene.

XX Streptococcus agalactiae.

XX Key

XX Location/Qualifiers

XX 1..1509

XX /tag= a

XX /product= "SbpI"


```

QY 1501 CGTCTTAA 1509
Db 1501 CGTCTTAA 1509

RESULT 3
AA20108
ID AAX20108 standard; DNA; 1983 BP.
XX
AC AAX20108;
XX
DT 20-APR-1999 (first entry)
XX
DE Enterococcus faecalis gene EF058.
XX
KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW detection; attenuation; antigenic; ss.
XX
OS Enterococcus faecalis.
XX
PN WO980554-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US008959.
XX
PR 06-MAY-1997; 97US-0044031P.
PR 16-MAY-1997; 97US-0046655P.
PR 14-NOV-1997; 97US-0066009P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kunsch CA, Choi GH, Bailey C, Hromockyj A;
XX
WPI: 1999-070095/06.
P-PSDB; AAY00118.
XX
New isolated Enterococcus faecalis polynucleotides - used to develop
products for the detection of Enterococcus and for use in vaccines for
prevention or attenuation of Enterococcus infection.
XX
Claim 1; Page 133; 301pp; English.
XX
The present sequence represents a gene isolated from Enterococcus
faecalis. The present invention describes genes, proteins and antigenic
polypeptides isolated from E. faecalis. The proteins can be used in
vaccines for preventing or attenuating an infection caused by a member of
the Enterococcus genus in an animal. They can also be used for detecting
Enterococcus antibodies in a sample. The nucleotide sequences can be used
for detecting Enterococcus nucleic acids. Products from the present
invention can also be used for screening compounds to identify agonists
and antagonists of E. faecalis protein activity
XX
Sequence 1983 BP; 729 A; 330 C; 426 G; 498 T; 0 U; 0 Other;

Query Match 4.3%; Score 64.8; DB 2; Length 1983;
Best Local Similarity 52.1%; Pred. No. 0.00013;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAGCTAATGCAACAGATATACACAGGACAGATGGTATATACATTACAGGC 1239
Db 1660 ACAAAAGCTGAAGCAACTACTTTTACAAACACGCGCTGATGGATTAGTATATCAGAGG 1719
QY 1240 TTGAAGAAGGTACATCTACTATCTAGTTGAGAAAAGGCTCCCTTAGGTTTCAATTTGTTA 1299
Db 1720 CTTAAATACGGTACTATATTAGAGAAAGTGTAGTCTCTGATGATGATGTTGTTA 1779
QY 1300 GATHAATCTCAGAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 1780 ACAAAATCGGATGAATTTGGTGCATGAAATCAATCATATGCGCAAA-----CAGAAAC 1833
QY 1360 CTTTATTAAACCAACTGTTGAAATAACAAAGGTAATGATGTTGCCCTTCAACAGGTGGT 1419

Db 1834 CTAGTTTCACAGAAAAAGTACCAACAAACAAAGGTACCTTTACCTTCAACAGGTGGC 1893
QY 1420 ATTGGTACAAATTTTCTACATTATAGGTGCAATTTTAGTATATAGGACGAGGTATCGTG 1479
Db 1894 AAAGGAATCTACGTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTGCGAGGAGTCTAC 1953
QY 1480 CTTGTTGCTCGTCTGCTGTTTACGTTTCTTAA 1509
Db 1954 TTGCTAGACGTAGAAAAGAAATGCTTAA 1983

RESULT 4
ABN98093
ID ABN98093 standard; DNA; 1983 BP.
XX
AC ABN98093;
XX
DT 05-AUG-2002 (first entry)
XX
DE E faecalis EF058 gene.
XX
KW Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;
KW gene; ds.
XX
OS Enterococcus faecalis.
XX
PN US2002045737-A1.
XX
PD 18-APR-2002.
XX
PF 04-MAY-1998; 98US-00071035.
XX
PR 04-MAY-1998; 98US-00071035.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX
WPI: 2002-425450/45.
P-PSDB; ABP43337.
XX
New genes and polypeptides from Enterococcus faecalis, useful as vaccines
for preventing, treating or attenuating an infection caused by a member
of the Enterococcus genus in an animal, particularly E. faecalis.
XX
Claim 1; Page 92; 255pp; English.
XX
The present invention provides the protein and coding sequences of a
number of polypeptides from Enterococcus faecalis. The proteins can be
used as vaccines for preventing or attenuating an infection caused by a
member of the Enterococcus genus in an animal, particularly E. faecalis.
XX
The polynucleotide is also useful for preventing or treating E. faecalis
infection. The present sequence is a coding sequence of the invention
XX
Sequence 1983 BP; 729 A; 330 C; 426 G; 498 T; 0 U; 0 Other;

Query Match 4.3%; Score 64.8; DB 6; Length 1983;
Best Local Similarity 52.1%; Pred. No. 0.00013;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAGCTAATGCAACAGATATACACAGGACAGATGGTATATACATTACAGGC 1239
Db 1660 ACAAAAGCTGAAGCAACTACTTTTACAAACACGCGCTGATGGATTAGTATATCAGAGG 1719
QY 1240 TTGAAGAAGGTACATCTACTATCTAGTTGAGAAAAGGCTCCCTTAGGTTTCAATTTGTTA 1299
Db 1720 CTTAAATACGGTACTATATTAGAGAAAGTGTAGTCTCTGATGATGATGTTGTTA 1779
QY 1300 GATHAATCTCAGAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 1780 ACAAAATCGGATGAATTTGGTGCATGAAATCAATCATATGCGCAAA-----CAGAAAC 1833
QY 1360 CTTTATTAAACCAACTGTTGAAATAACAAAGGTAATGATGTTGCCCTTCAACAGGTGGT 1419

```


QY 1360 CTTTGTAGTAAACCACTGTTGAAATACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419
 DB 1834 CTAGTTTCACAGAAAAGTACCAACAAACAAACAAAGGTACTGAGTTGCTTCAACAGGTGGC 1893
 QY 1420 ATTGGTACAAACATTTTCTACATTATAGTGTGCAATTTTAGTAATAGGAGCAGGTATCGTG 1479
 DB 1894 AAAGGAATCTACGTTTACTTAGGAGTGGCGAGTCTTGCTACTTATTGCGAGGAGTCTAC 1953
 QY 1480 CTTGTTGCTGCTGCTGCTGCTTACCTTTCTTAA 1509
 DB 1954 TTTGCTAGACGTAGAAAAGAAAATGCTTAA 1983

RESULT 5

ACA88057
 ID ACA88057 standard; DNA; 1983 BP.

XX AC ACA88057;

DT 07-JUL-2003 (first entry)

XX E. faecalis novel gene #109.

XX Gene; ds; endocarditis; bacteraemia; urinary tract infection; UTI;
 KW intraabdominal infection; soft tissue infection; neonatal sepsis;
 KW vaccine.

XX Enterococcus faecalis.

XX US2003017495-A1.

XX 23-JAN-2003.

XX 29-JUL-2002; 2002US-00206576.

XX 06-MAY-1997; 97US-0044031P.

XX 16-MAY-1997; 97US-0046655P.

XX 14-NOV-1997; 97US-0066009P.

XX 04-MAY-1998; 98US-00071035.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;

XX WPI; 2003-416890/39.

XX P-PSDB; ABU88365.

XX New nucleic acid molecules and polypeptides from Enterococcus faecalis,
 PT useful as vaccines for preventing or attenuating an enterococcal
 PT infection in an animal, or for identifying Enterococcus faecalis in
 PT biological samples.

XX Claim 1; Page; 40pp; English.

XX The invention relates to a new isolated nucleic acid molecule comprising
 CC a polynucleotide isolated from Enterococcus faecalis appearing as
 CC ACA87949-ACA88196 (for sequences complementary to them or 95% identical to
 CC them). Also included are the proteins encoded by the above nucleic acids,
 CC making a recombinant vector (comprising inserting the isolated nucleic
 CC acid molecule cited above into a vector), a host cell comprising the
 CC vector (used to produce the protein), an isolated antibody specific for
 CC the polypeptides, a hybridoma that produces the antibody, an isolated
 CC polypeptide antigen comprising an amino acid sequence of an Enterococcus
 CC faecalis epitope listed in the specification, a vaccine comprising one or
 CC more E. faecalis polypeptides (and a pharmaceutical diluent, carrier or
 CC excipient) where the polypeptide elicits protective antibodies in an
 CC animal to a member of the genus Enterococcus; preventing or attenuating
 CC an infection caused by a member of the genus Enterococcus in an animal
 CC comprising administering to the animal the polypeptide and detecting
 CC Enterococcus nucleic acids in a biological sample. The E. faecalis
 CC nucleic acid molecules and polypeptides are useful as vaccines for
 CC preventing or attenuating an enterococcal infection in an animal (e.g.
 CC endocarditis, bacteraemia, urinary tract infection (UTI), intraabdominal

CC infection, soft tissue infection and neonatal sepsis). The polypeptides
 CC are also useful for detecting Enterococcus aureus in immunoassays, as
 CC epitope tags, as molecular weight markers, or for generating antibodies
 CC that specifically bind E. faecalis polypeptides. The nucleic acid
 CC molecules are also useful as probes for gene mapping, or for identifying
 CC E. faecalis in biological samples. The kit and methods are useful for
 CC detecting Enterococcus antibodies or nucleic acid molecules in a
 CC biological sample. The present sequence is a novel E. faecalis nucleic
 CC acid of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from the USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030017495

XX Sequence 1983 BP; 729 A; 330 C; 426 G; 498 T; 0 U; 0 Other;

Query Match 4.3%; Score 64.8; DB 7; Length 1983;

Best Local Similarity 52.1%; Pred. No. 0.00013;

Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAGCTAATGCAACAGATATACACAGGAGCAGATGGTATATACCATTCACAGGC 1239

DB 1660 ACAAAAGCTGAAGCAACTACTTTTACACAAACGCGCTGATGGATTGTTGATATCACAGGG 1719

QY 1240 TTGAAAGAGGTACATATCTATCTAGTTGAGAAAAAGGCTCCCTTAGGTTACAATTTGTTA 1299

DB 1720 CTTAATACGGTACCTATTATTAGAGAACTGTAGCTCCTGATGATGTTGTTGTTA 1779

QY 1300 GATACTCTCAGAGGTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359

DB 1780 ACAATCGGATTGAATTTGTGGTCAATGAACATCATATGGCACA-----CAGAAAC 1833

QY 1360 CTTTGTAGTTAACCACTGTTGAAATACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419

DB 1834 CTAGTTTCACAGAAAAAGTACCAAAACACAAAGGTACCTTACCTTCAACAGGTGGC 1893

QY 1420 ATTGGTACACAAATTTTCTACATTATAGTGTGCAATTTTAGTAATAGGAGCAGGTATCGTG 1479

DB 1894 AAAGGAATCTACGTTTACTTAGGAAGTGGCGAGTCTTGCTACTATTGCGAGGTCATC 1953

QY 1480 CTTGTTGCTGCTGCTGCTGCTTACGTTCTTAA 1509

DB 1954 TTTGCTAGACGTAGAAAAAGAAAATGCTTAA 1983

RESULT 6

ABX61663

ID ABX61663 standard; DNA; 1983 BP.

XX AC ABX61663;

XX DT 26-FEB-2003 (first entry)

XX Enterococcus faecalis EF040 polynucleotide #109.

XX EF040; gene; ds; immunostimulant; antibacterial; gene mapping.

XX Enterococcus faecalis.

XX US6448043-B1.

XX 10-SEP-2002.

XX 04-MAY-1998; 98US-00071035.

XX 06-MAY-1997; 97US-0044031P.

XX 16-MAY-1997; 97US-0046655P.

XX 14-NOV-1997; 97US-0066009P.

XX 14-NOV-1997; 97US-0066009P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;

XX WPI; 2003-416890/39.

XX P-PSDB; ABU88365.

XX New nucleic acid molecules and polypeptides from Enterococcus faecalis,
 PT useful as vaccines for preventing or attenuating an enterococcal
 PT infection in an animal, or for identifying Enterococcus faecalis in
 PT biological samples.

XX Claim 1; Page; 40pp; English.

XX The invention relates to a new isolated nucleic acid molecule comprising
 CC a polynucleotide isolated from Enterococcus faecalis appearing as
 CC ACA87949-ACA88196 (for sequences complementary to them or 95% identical to
 CC them). Also included are the proteins encoded by the above nucleic acids,
 CC making a recombinant vector (comprising inserting the isolated nucleic
 CC acid molecule cited above into a vector), a host cell comprising the
 CC vector (used to produce the protein), an isolated antibody specific for
 CC the polypeptides, a hybridoma that produces the antibody, an isolated
 CC polypeptide antigen comprising an amino acid sequence of an Enterococcus
 CC faecalis epitope listed in the specification, a vaccine comprising one or
 CC more E. faecalis polypeptides (and a pharmaceutical diluent, carrier or
 CC excipient) where the polypeptide elicits protective antibodies in an
 CC animal to a member of the genus Enterococcus; preventing or attenuating
 CC an infection caused by a member of the genus Enterococcus in an animal
 CC comprising administering to the animal the polypeptide and detecting
 CC Enterococcus nucleic acids in a biological sample. The E. faecalis
 CC nucleic acid molecules and polypeptides are useful as vaccines for
 CC preventing or attenuating an enterococcal infection in an animal (e.g.
 CC endocarditis, bacteraemia, urinary tract infection (UTI), intraabdominal

Db 13844 ACAATCGGATTGAATTTTGGTCAATGAACAATCATATGGCACAA-----CAGAAAC 13897
QY 1360 CTTTGTAGTTAACCCAACTGTTGAAATAACAAAGGTACTGAGTTCCTTCAACAGGTGGT 1419
Db 13898 CTAGTTTACACAGAAAGTACCAACAAACACAAAGGTACTTACCTTCAACAGGTGGC 13957
QY 1420 ATTGTACACAAATTTCTACATTTATAGGTGCAATTTTATAGTAAGGACAGGTATCGTG 1479
Db 13958 AAAGGAATCTAGTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTGACAGGAGTCTAC 14017
QY 1480 CTTGTTGCTCGTGGTTCAGTTCTTTAA 1509
Db 14018 TTGCTAGACGTAGAAAGAAATGCTTAA 14047

RESULT 10

AAA70187
ID AAA70187 standard; DNA; 11922 BP.
XX AC
XX AAA70187;
XX DT
XX 07-NOV-2000 (first entry)
XX DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:320.
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX OS Plasmodium falciparum.

W0200025728-A2.

11-MAY-2000.

05-NOV-1999; 99WO-US026796.

05-NOV-1998; 98US-0107131P.

(HOFF// HOFFMAN S.

(CARU// CARUCCI D.

(GARD// GARDNER M.

(VENT// VENTER J C.

Hoffman S, Carucci D, Gardner M, Venter JC;

WPI; 2000-365347/31.

Proteins encoded by chromosome 2 of the human malarial parasite,
Plasmodium falciparum, useful as antimalarial vaccines and in the
diagnosis of P.falciparum infection.

Disclosure; Page 516-519; 577pp; English.

The present invention describes proteins and their fragments (I) encoded
by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
vaccines against P. falciparum infection comprising (I) or (II). (I) and
(II) are useful for the development of vaccines against P. falciparum
infection. (I) and polyclonal antisera or a monoclonal antibody raised to
immunogens comprising the sequences of (I), are useful in the detection
of infection with P. falciparum. Furthermore, (I) (especially when they
are refined or secreted or membrane proteins) can aid the identification
of drugs to treat or prevent P. falciparum infection, or they can be used
to identify drug resistance in P. falciparum. Sequencing of the
Plasmodium chromosome 2 and the subsequent identification of proteins
encoded by it will help to expand our understanding of parasite biology,
a process hampered by the complexity of the parasitic lifecycle, and
provide new targets for vaccine and drug development. Parasite resistance
to drugs and mosquito resistance to insecticides have led to a resurgence
of malaria in many parts of the world, and there is a pressing need for
CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
represent nucleotide and protein sequences given in the present

CC invention, but which are not specifically mentioned within the
CC specification
XX SQ Sequence 11922 BP; 5402 A; 948 C; 1343 G; 4229 T; 0 U; 0 Other;
Query Match 3.9%; Score 59.6; DB 3; Length 11922;
Best Local Similarity 46.3%; Pred. No. 0.0029;
Matches 265; Conservative 0; Mismatches 304; Indels 3; Gaps 2;
QY 111 TCAGATACCTCAAAAGGCGCAACCTATAAAGCATATAAAGTTTTTGTATCGAGAAATAGA 170
Db 10038 TCTAGATATTATTAATGACCATACCTTTATGGAGTCTATGAAAGTTGCAACCAATAGGAT 10097
QY 171 TAATGCAATGTATCTGATTGGAATAAAGATGAGCTTCTTATTTTAACTTCTCAAGGTAA 230
Db 10098 TTATGAAATCCATCTACTATAGGAATTTATATAAAGAACCGTACACCTCTTAATGAACAA 10157
QY 231 AGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATCTCTTTTTACGCAACTACTAATGG 290
Db 10158 AATTTTAAATAGTAAATAGTATTAACAATAGTAAATCATTTTAAAGAAAGTAAATGATGA 10217
QY 291 AGGAGAACATATGTAACCTAAAAAGATCTGCTGAGCAATGAGATTCGCGACATGGGC 350
Db 10218 AAATGTAAGAAACAAACACAAATACAAATACAAATACAAATACAAATACAAATAA 10277
QY 351 TAAATCTATATCAGCTAATAACTACACAGTTTCCACTGTTTACTGAGTCAAAATTAATGATGG 410
Db 10278 AATTTCTGATACAGATAATCAATATGATGCAATATAATCAATGATGCAATATAATCAATA 10337
QY 411 TACTGAGGTTTAAATGTTTCCCAATATGAGTATTAATGTTTCT-AGCACTGTTTAATA 469
Db 10338 TGATGATATAATCATAATGACGATATAATCATAATGATGATGATATAATCATAATGATGC 10397
QY 470 ATGGAGCTGTAATTATGTTTACATCTGTAA--CTCCAAATGCTACTATTTCATGAGAAAGA 527
Db 10398 ATATAATCATAATGATGATGATATAATCATAATGATGATGATATAATCATAATGATGATATA 10457
QY 528 TACTGATGCGACATGGGAGATGGTGGTGAAGAACTGTAGATCAAAAAACGCTACTCGGT 587
Db 10458 TCATAATGATGCAGATAATCATAATGATACAGATAATCATAGTATGATTAATTAATTTCTCA 10517
QY 588 TGGTATACAGTCAAAATATACTATTACTTTATTAAGATGAGTCAATATATCATGTTACAGA 647
Db 10518 TAAATATAAGGAACTTATAAAATTTATCGTATTTCATGATGAGGAAGATATTATTCAAGA 10577
QY 648 AAAAGTGATCAATATGTTATAAAGGATACTA 679
Db 10578 TAATAATTATACAAACGATGATTGTTAATA 10609

RESULT 11

ACA28118
ID ACA28118 standard; DNA; 1137 BP.

XX AC ACA28118;

XX DT 19-JUN-2003 (first entry)

XX DE Prokaryotic essential gene #9775.

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;

XX KW drug design; gene.

XX OS Clostridium acetobutylicum.

XX PN W0200277183-A2.

XX XX 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX XX 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

```
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (BLIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX P-PSDB; ABU24248.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 15988; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is one of the target
XX prokaryotic essential genes. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1137 BP; 450 A; 138 C; 217 G; 332 T; 0 U; 0 Other;
XX
XX Query Match 3.9%; Score 59.2; DB 7; Length 1137;
XX Best Local Similarity 43.7%; Pred. No. 0.0021;
XX Matches 262; Conservative 0; Mismatches 338; Indels 0; Gaps 0;
XX
XX 90 TGAGACAGGACAAATACAGTTCAAGTACTCAAAAGGCGCAACCTATAAGCATATA 149
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 30 TGGGACAGGCCAGAGCTATAAGATGGCTCTCTTTGTTAAATACTGAAGAAATA 89
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 150 AGTTTGTGTCAGAAATAGATATGCAATGTATCTGATTCGAATAAGATGAGCTTC 209
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 90 AAATTTTCAAGAAAAGTAGTGTGTACAGCAGCAGGAGATGCTAGATCAAGTTCT 149
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 210 TTATTTTAATTCCTCAAGTAAAGAGCTGAGTATAAGCTTCAACTGATTTTAATCTCT 269
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 150 TAGTTTGTTCGAATAGAGCCAGCTTTGATTTGATATATGAATAAAGCAATCTCT 209
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 270 TTTTACAGCACTACTAATGAGGGGAGAAATATGTAACTTAAAGAAATAGTCTGCTCAGC 329
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 210 TACACAGTAACCTCAAGTCTTTTACATGGGTTAAGCGAATATTTCTGCTGAAGACC 269
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 330 AAATGAGATTGCGACATGGGCTAAATCTATATCAGCTAATATACACAGTTTCCACTGT 389
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX Db 270 AGATATATTTAGTACATGGAGATACAACTACAGTTTTCGAGCATCTTCGAGCCTT 329
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Qy 390 TACTGAGTCAAATAATGATGGTACTGAGGTATTAATGTTTCCCAATATGATATATTA 449
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 330 TTATGAAAGATAGACTATTGGGCATGTTGAGGCTGGACTTAGAACTTATGATAAATATT 389
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Qy 450 TGTTCCTAGCAGCTGTTAATAATGAGAGCTGTAATATGTTTACATCTGTAACTCCAATGC 509
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 390 TCTTTTCTCGAAGAAAGTAAATAGAAAACCTTACTGGAGCAATTCAGATATGCAATTTCG 449
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Qy 510 TACTATTCATGAAAGAAATACTATCGCATCGGAGATGGTGGTGGAAAAAATCTGTAGA 569
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 450 ACCGACAGTAAATCTTAAAAAATAATCTTTTAGAGAGAGGTGTAAAGAAAAAATATATT 509
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Qy 570 TCAAAAAACGTTACTCGGTTGGTGATACAGTCAGTCAAAATATCTATTACTTTATAAGAAATGCAGT 629
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 510 TATAACTGGAAAATCTGCTGATTGCTATGAAATACACAGTCGATGTAATATGATGATT 569
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Qy 630 CAATTCATCGGTACAGAAAAAGTGTCAATATGTTATAAGAGATATGCACTGCTGC 689
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 570 TAAAAATGATCAATTAATAAATTTGAACTACAAAACATAAAAGATATTATTATGTTAACTGC 629
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX RESULT 12
XX ABQ39490/C
XX ID ABQ39490 standard; DNA; 969 BP.
XX AC ABQ39490;
XX XX
XX DT 12-JUL-2002 (first entry)
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 26081.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX Homo sapiens.
XX WO200218632-A2.
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP010074.
XX
XX 01-SEP-2000; 2000DB-01043826.
XX
XX 05-SEP-2000; 2000DB-01044543.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used:
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
```

Db 9 TAATAAA 3
| | | |
RESULT 13
ABQ39491
ID ABQ39491 standard; DNA; 969 BP.
XX
XX
AC ABQ39491;
XX
DT 12-JUL-2002 (first entry)
XX
DE
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 26082.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26082.
 XX
 XX
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism
 KW SNP; cell differentiation; de.
 XX
 XX Homo sapiens.
 OS
 XX WO200218632-A2.
 XX
 XX 07-MAR-2002.
 XX
 XX 01-SEP-2001; 2001WO-EP010074.
 XX
 XX 01-SEP-2000; 2000DE-01043826.
 XX
 XX 05-SEP-2000; 2000DE-01044543.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;
 PI
 XX WPI; 2002-371829/40.
 XX
 XX
 XX Determining the degree of cytosine methylation in genomic DNA, useful
 PT diagnosis and prognosis, comprises selective hybridization of amplicon
 PT from chemically treated DNA.
 XX
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label
 CC the amplicon. From the ratio of labels hybridised to the two classes
 CC oligomers, the degree of methylation is calculated. The method is used
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc
 CC particularly by detecting mutations or single nucleotide polymorphism
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention

SQ	Sequence	969'BP;	645 A;	93 C;	54 G;	177 T;	0 U;	0 Other;	
	Query Match		3.8%;	Score	56.6;	DB	6;	Length	969;
	Best Local Similarity		42.5%;	Pred. No.	0.0075;				
	Matches	360;	Conservative	0;	Mismatches	484;	Indels	3;	Gaps
QY	90	TGACACAGGACAAATTACAGTTCACAGATAC	CTCAAAAGGGCGCAACCTATAAAGCATATAA	149					
Db	121	TAAAAACGAATAAAAAATACGAATAAAAAATCTATAAAAAACGAATAAAAAATAAAAATAAAAA		180					

XX AmEPV genome fragment#5.
 XX AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
 KW genetic deficiency disorder; ds.
 XX Amsacta moorei entomopoxvirus.
 OS WO200212526-A2.
 XX 14-FEB-2002.
 XX 10-AUG-2001; 2001WO-US025287.
 XX 10-AUG-2000; 2000US-0224479P.
 PR 14-SEP-2000; 2000US-00662254.
 XX (UYFL) UNIV FLORIDA.
 XX Moyer RW, Li Y, Bawden AL;
 PI WPI; 2002-227161/28.
 DR Novel recombinant entomopox virus vector useful for delivering
 PT polynucleotide encoding protein to vertebrate cell, comprises
 PT polynucleotide encoding protein operably linked with heterologous
 PT promoter sequence.
 XX Disclosure; Page 226-242; 326pp; English.
 XX The invention relates to a recombinant entomopox virus (EPV) vector,
 CC comprising a polynucleotide encoding a protein operably linked with a
 CC heterologous promoter sequence. The invention also concerns methods for
 CC providing gene therapy for genetic deficiency disorders. Vectors of the
 CC invention are useful for delivering a polynucleotide encoding a protein
 CC to a vertebrate cell preferably a mammalian cell, such as a human cell.
 CC The vector is introduced into the vertebrate cell by infection in a viral
 CC particle, or by transfection, transduction, or injection either in vitro
 CC or in vivo. The vector is useful for the delivery and expression of
 CC biologically useful proteins in gene therapy protocols, and for
 CC delivering large DNA segments for engineering of vertebrate cells.
 CC Polynucleotides of the invention have applications in techniques such as
 CC their use as insertion sites for foreign genes of interest, hybridisation
 CC probes, for chromosome and gene mapping, in PCR technologies, and in the
 CC production of sense or antisense nucleic acids. Vectors of the invention
 CC provide for stable integration and expression of heterologous DNA in host
 CC cells, and are adapted for accepting large heterologous polynucleotide
 CC inserts which can be delivered in an infected or transformed cell and
 CC expressed in a stable fraction. The current sequence represents a
 CC fragment of the genome of the genus B entomopoxvirus from amsacta moorei
 CC (AmEPV)
 XX
 SQ Sequence 32392 BP; 13748 A; 2577 C; 2550 G; 13517 T; 0 U; 0 Other;
 Query Match 3.7%; Score 55.2; DB 6; Length 32392;
 Best Local Similarity 44.0%; Pred. No. 0.034;
 Matches 329; Conservative 0; Mismatches 413; Indels 6; Gaps 2;
 QY 479 TAATTATGGTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAGAAATACGTGCGA 538
 DB 1602 TAATAATATAAATGGATATATTACTAATTCGCTCAACATCATCATATTTTCAGATCATA 1543
 QY 539 CATGGGAGATGGTGGGAAAGAACTGTAGATCAAAAAGCTACTCGTTGGTGATACAG 598
 DB 1542 ACATATATGTTGATGAGAGAAAATAGAGATTTTAATAATAATATATCTCGGTTTAA 1483
 QY 599 TCAATATATCTATTATTAAGAATGCGAGTCAATTATCATGGTACAGAAAAGTGTATC 658
 DB 1482 TTAATAATATATTA---ACAATAATACTAAAAAAAATATACTACAATAATAGATAATT 1426
 QY 659 AATATGTTATAAGGATACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 718
 DB 1425 ATAATGAATAGAAAAAAATAAATTAATGAATAAGTGTGGATAATAG---TG 1369

QY 719 ATGAAGTAACCTATTACTGATGGATCAGGAATATTACAACCTCAACTCAAGTTTCGAAA 778
 DB 1368 ATTGCACAGATCTTAACCTAATAAATAAGCAATATTATTCAAAATCAGTATATAGTAAA 1309
 QY 779 AAGCAACTGGGAAGTAGTAACCTGTTAGAGGAAAATAATAATTTTCAAGATTACTATCCGT 838
 DB 1308 CAGAAGAAGGGAGTAAATCTGTTATTAACTGATATCTGATAATAATTAATAAATAGTT 1249
 QY 839 GGGCAGCTACCAATATCTCCAAACCGGAAATACTCAAAATGGAGGCTAATGACTTTTTTT 898
 DB 1248 TAGCAGAAGATATATCTTTTAAAAAGAGGATGTATAAAGAAAGATTATAATATAGATTAT 1189
 QY 899 ATAAGGGAATAATAACATCAACAGTCACATTTATACAGAGCTATTAAAGAGTGGAGCTAAAC 958
 DB 1188 ATGAGAAAATAATAATGTAATATACAAATAGAAATAAATAATATTATATCAAAAAATTAT 1129
 QY 959 CAGGTTGAGCTGATTTTACCAGAAAATAACAACATTCGACCATCAACCCCAATACTAGCA 1018
 DB 1128 TATCTTTTGATTATAATAATAAAAAATAAATAATAATAGTGTGCTGTAATAATTGATGAA 1069
 QY 1019 ATGATGACCCAGTCAAAAAGTAAACAGTGGAGGATCGTCAATTAATCTATAAAAAATTG 1078
 DB 1068 TTATTACGATGATTAAATTTGAAAGCAAAAATCGAAGAAAAGAGACTATTCAATACATTC 1009
 QY 1079 ATGGTTCCCAAAAAGCTTCATTACAAGGTGCTATATTGTTTTTAAAGAATGCTACGGGTC 1138
 DB 1008 CTATATATGAAAAAGTTCAATTAGAAATATATTATTTTATTGATTAATAAAGAGTA 949
 QY 1139 AATTTCTAAACTTTAACGATACAAATAACGTTGAATGGGGCACAGAAGCTAATGCAACAG 1198
 DB 948 AATTAATTCACACTATATGATAACGAAATAGATGTATAGATTATATATACCGACGAACAAT 889
 QY 1199 AATATACAACAGGAGCAGATGGTATAAT 1226
 DB 888 TATTTCTGAATAATGTAATAATAATAAT 861

Search completed: September 13, 2004, 02:13:17
 Job time : 608.354 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 01:03:46 ; Search time 5983.67 Seconds
(without alignments)
10930.517 Million cell updates/sec

Title: US-10-009-254-1
Perfect score: 1509
Sequence: 1 atgaaaagaagaatgattca.....gtcgtcggttaacgtttcttaa 1509

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_inu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1509	100.0	1509	1	AF485279	AF485279 Streptococcus
2	1509	100.0	1509	6	AX476887	AX476887 Sequence
3	92.8	6.1	216959	10	AC116998	AC116998 Mus muscu
4	89.6	5.9	171104	2	BX649545	BX649545 Danio rer
5	87.2	5.8	170221	10	AC127421	AC127421 Mus muscu
6	86.4	5.7	213544	2	BX510939	BX510939 Danio rer
c 7	86	5.7	116807	2	BX890614	BX890614 Danio rer
c 8	85.2	5.6	146275	2	AL935272	AL935272 Danio rer
c 9	84	5.6	198743	2	BX530070	BX530070 Danio rer
c 10	83.4	5.5	155204	2	AC007926	AC007926 Trypanoso
c 11	82.4	5.5	94534	5	AL929250	AL929250 Zebrafish
c 12	80.6	5.3	110000	2	PFMAL13.07	Continuation (8 of
c 13	80.4	5.3	225581	2	BX537105	BX537105 Danio rer
c 14	80	5.3	10115	5	AF397467	AF397467 Ictalurus
c 15	79.8	5.3	308092	3	AY333070	AY333070 Drosophila
c 16	78.6	5.2	224391	2	AC137124	AC137124 Mus muscu
c 17	78	5.2	164399	3	PFMAL3P6	298551 Plasmodium
c 18	77.8	5.2	154160	2	BX571709	BX571709 Danio rer
c 19	77.8	5.2	177595	2	AL953867	AL953867 Danio rer
c 20	77.4	5.1	200634	5	BX004990	BX004990 Zebrafish
c 21	77.2	5.1	176010	5	AL935062	AL935062 Zebrafish
c 22	75.6	5.0	151341	5	AL929536	AL929536 Zebrafish
c 23	75.4	5.0	22398	5	FRU271723	AL9271723 Fugu rubr
c 24	75.2	5.0	245252	2	BX088589	BX088589 Danio rer
c 25	75	5.0	183357	2	BX569779	BX569779 Danio rer
c 26	75	5.0	183357	2	BX569779	BX569779 Danio rer
c 27	74	4.9	199698	2	BX004991	BX004991 Danio rer
c 28	73.8	4.9	250029	3	AB014820	AB014820 Plasmodiu
c 29	73	4.8	666	8	AF413050S2	AF413051 Zea mays
c 30	73	4.8	199347	2	BX321884	BX321884 Danio rer
c 31	72.8	4.8	146907	2	AC141031	AC141031 Rattus no
c 32	72.6	4.8	173134	5	AL935282	AL935282 Zebrafish
c 33	72.4	4.8	157033	2	BX323881	BX323881 Danio rer
c 34	72.4	4.8	179553	2	AC024253	AC024253 Homo sapi
c 35	72.2	4.8	153751	3	AC116551	AC116551 Dictyoste
c 36	72.2	4.8	180668	2	AC020857	AC020857 Mus muscu
c 37	70.6	4.7	146415	5	AL929469	AL929469 Zebrafish
c 38	70.6	4.7	212722	2	AC098909	AC098909 Rattus no
c 39	70.2	4.7	2167	3	PRSTARPA	Z30339 P. reichenow
c 40	69.6	4.6	202182	2	AC120639	AC120639 Rattus no
c 41	69.6	4.6	274060	2	AC095184	AC095184 Rattus no
c 42	69.4	4.6	203271	2	BX322566	BX322566 Danio rer
c 43	69.2	4.6	11534	3	PFAARPIPR	Y08926 P. faiciparu
c 44	69	4.6	178664	2	BX511115	BX511115 Danio rer
c 45	69	4.6	180023	2	BX897685	BX897685 Danio rer

ALIGNMENTS

RESULT 1
AF485279 1509 bp DNA linear BCT 25-NOV-2003
LOCUS Streptococcus agalactiae Spbl (spbl) gene, complete cds.
DEFINITION Streptococcus agalactiae
ACCESSION AF485279
VERSION AF485279.1 GI:28628942
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1 (bases 1 to 1509)
AUTHORS Adderson,E.E., Takahashi,S., Wang,Y., Armstrong,J., Miller,D.V. and
Bohnsack,J.F.
TITLE Subtractive Hybridization Identifies a Novel Predicted Protein

Mediating Epithelial Cell Invasion by Virulent Serotype III Group B
Streptococcus agalactiae
 Infect. Immun. 71 (12), 6857-6863 (2003)
 14638773
 2 (bases 1 to 1509)
 Adderson, E. and Bohnsack, J. F.
 Direct Submission
 Submitted (19-FEB-2002) Infectious Diseases, St. Jude Children's
 Research Hospital, 332 N. Lauderdale St., Memphis, TN 38105, USA
 Location/Qualifiers
 1..1509
 /organism="Streptococcus agalactiae"
 /mol_type="genomic DNA"
 /strain="874391"
 /serotype="RDP type III-3"
 /db_xref="taxon:1311"
 1..1509
 /gene="spbl"
 1..1509
 /gene="spbl"
 /note="adhesin/invasin; surface protein of group B
 streptococcus 1"
 /codon_start=1
 /transl_table=11
 /product="Spbl"
 /protein_id="AAO49409.1"
 /db_xref="GI:28628943"
 /translation="MKKMIQSLVASLFGMAVSPVPIAFRAETGTTIVQDPOKA
 TYKAVFDSIDNANSDSKNGASVLIPOGKEAEYKASDSDNSLFTTYTINGGRIV
 TKDTSANEIATWAKISANTPVSITSNNDGTEVINVSQYIYVSSVNVNGAV
 IMVTSVPNTAHSKNDATWGDGGKTVQDQYSVGTIVKYITTYKNAVNYHGTKEV
 YOYVKOTMPSAVDLNDESVETITDGSNIITLTQSEKATGKNLLEENNFTI
 TIPAANTPTQNGANDDFKGINITITVTVGLKSGAKPGSADLPENTNIAT
 NPNTSNDPQKVTVRDQITIKKIDGSTRASLOQAFVLKNAFGQFLNFDNTNVEM
 GTEANETVYTGAGLITIGLKEGTLYLVEKAPLGVNLLDSQKVLGDGADFTN
 SDNLLNPTVNNKGTLPSTGGIGTTFIYIIGALLVIGAGIVLVARRLRIS"

FEATURES

source

gene

CDS

ORIGIN

Query Match 100.0%; Score 1509; DB 1; Length 1509;
 Best Local Similarity 100.0%; Pred. No. 3e-238;
 Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAAGAAATGATTCAATCGCTGTAGTGGCGAGTTTAGCAATTTGGTATGGCTGTA 60
 DB 1 ATGAAAAGAAATGATTCAATCGCTGTAGTGGCGAGTTTAGCAATTTGGTATGGCTGTA 60

QY 61 TCACCAAGTACCGCGATAGCTTTTCCGCGTCGAGACAGGCAATTTACAGTTCAAGTACT 120
 DB 61 TCACCAAGTACCGCGATAGCTTTTCCGCGTCGAGACAGGCAATTTACAGTTCAAGTACT 120

QY 121 CAAAAGCGCAACTATAAGCATAAAGTTTTCATGCAGAAATAGATAATGCAAAAT 180
 DB 121 CAAAAGCGCAACTATAAGCATAAAGTTTTCATGCAGAAATAGATAATGCAAAAT 180

QY 181 GTATCTGATTCGAATAAGATGGAGCTTCTATTATTTCTTCTTACGCACTACTAATGAGAGCTGAG 240
 DB 181 GTATCTGATTCGAATAAGATGGAGCTTCTATTATTTCTTCTTACGCACTACTAATGAGAGCTGAG 240

QY 241 TATAAAGCTTCAACTGATTTTAAATTTCTTTTACGCACTACTAATGAGAGGAGACA 300
 DB 241 TATAAAGCTTCAACTGATTTTAAATTTCTTTTACGCACTACTAATGAGAGGAGACA 300

QY 301 TATGTAATAAAGATAGTGGCTGACAAATAGAGATGCGACATGGGCTAAATCTATA 360
 DB 301 TATGTAATAAAGATAGTGGCTGACAAATAGAGATGCGACATGGGCTAAATCTATA 360

QY 361 TCAGCTAATACTACACAGTTTCCAGTGTACTGAGTCAATATGATGCTAGAGTT 420
 DB 361 TCAGCTAATACTACACAGTTTCCAGTGTACTGAGTCAATATGATGCTAGAGTT 420

QY 421 ATTAATGTTTCCCAATATGATATTTATGTTTCTAGCACTGTTAATATGAGCTGTA 480
 DB 421 ATTAATGTTTCCCAATATGATATTTATGTTTCTAGCACTGTTAATATGAGCTGTA 480

QY 481 ATTATGTTACATCTGTAACTCCAAATGCTACTATTTCATGAAGAATACATGATCGACA 540
 DB 481 ATTATGTTACATCTGTAACTCCAAATGCTACTATTTCATGAAGAATACATGATCGACA 540

QY 541 TGGGAGATGGTGGGAAAACCTGTAGATCAAAAAACGTACTCGGTTGGTGATACAGTC 600
 DB 541 TGGGAGATGGTGGGAAAACCTGTAGATCAAAAAACGTACTCGGTTGGTGATACAGTC 600

QY 601 AAATATATCTATTCTTATATAAGATGCAATCAATATCATGGTACAGAAAAAGTGATCAA 660
 DB 601 AAATATATCTATTCTTATATAAGATGCAATCAATATCATGGTACAGAAAAAGTGATCAA 660

QY 661 TATGTTATAAAGGATACATGCGCATCTCTCTCTGTAGTGTGATTTGACGAGGCTCTAT 720
 DB 661 TATGTTATAAAGGATACATGCGCATCTCTCTCTGTAGTGTGATTTGACGAGGCTCTAT 720

QY 721 GAAGTAACTATTACTGTATGATGAGGGAATATTACAACCTCAACTCAAGGTTTCGAAAAA 780
 DB 721 GAAGTAACTATTACTGTATGATGAGGGAATATTACAACCTCAACTCAAGGTTTCGAAAAA 780

QY 781 GCAACTGGGAAGTATAACCTGTTAGAGGAAAAATAATAATTTACAGTACTATTCCGTGG 840
 DB 781 GCAACTGGGAAGTATAACCTGTTAGAGGAAAAATAATAATTTACAGTACTATTCCGTGG 840

QY 841 GCAGCTACCAATATCTCCAAACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTTTAT 900
 DB 841 GCAGCTACCAATATCTCCAAACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTTTAT 900

QY 901 AAGGGAATAAATAACAATCAAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960
 DB 901 AAGGGAATAAATAACAATCAAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960

QY 961 GGTTCAGCTGATTTACGAGAAATAAACAATTCGACCATCAACCCCAATACATAGCAAT 1020
 DB 961 GGTTCAGCTGATTTACGAGAAATAAACAATTCGACCATCAACCCCAATACATAGCAAT 1020

QY 1021 GATGACCCAGGTCAAAAAGTAAACAGTGAGGAGTGGTCAAAATTAATAAAAAAATTTGAT 1080
 DB 1021 GATGACCCAGGTCAAAAAGTAAACAGTGAGGAGTGGTCAAAATTAATAAAAAAATTTGAT 1080

QY 1081 GGTTCACAAAAGCTTCATTACAGGTGCTATATTTGTTTTTAAAGAAATGCTACGGGTCAA 1140
 DB 1081 GGTTCACAAAAGCTTCATTACAGGTGCTATATTTGTTTTTAAAGAAATGCTACGGGTCAA 1140

QY 1141 TTTCTAAACCTTTAACGATACAAATACGTTGAATGGGCGACAGAGCTAATGCAACAGAA 1200
 DB 1141 TTTCTAAACCTTTAACGATACAAATACGTTGAATGGGCGACAGAGCTAATGCAACAGAA 1200

QY 1201 TATACAACAGGACAGATGGTATTAATTAACCATTTACAGGCTTGAAAAGAGGTACATAT 1260
 DB 1201 TATACAACAGGACAGATGGTATTAATTAACCATTTACAGGCTTGAAAAGAGGTACATAT 1260

QY 1261 CTAGTTTGAGAAAAAGGCTCCCTTAGTTACAATTTGTTAGATAAATCTCAGAAAGGTTATT 1320
 DB 1261 CTAGTTTGAGAAAAAGGCTCCCTTAGTTACAATTTGTTAGATAAATCTCAGAAAGGTTATT 1320

QY 1321 TTAGGAGATGGAGCCACTGATACGCTAATTCAGATAACCTTTTGTAGTAAACCACTGTT 1380
 DB 1321 TTAGGAGATGGAGCCACTGATACGCTAATTCAGATAACCTTTTGTAGTAAACCACTGTT 1380

QY 1381 GAAAAATAACAAAGGTACTGAGTTGCTTCAACAGGTTGGTATTTGGTACAAATTTTCTAC 1440
 DB 1381 GAAAAATAACAAAGGTACTGAGTTGCTTCAACAGGTTGGTATTTGGTACAAATTTTCTAC 1440

QY 1441 ATTATAGTGGCAATTTTAGTAAATAGAGCAGGATTCGTCCTGTTGCTCGTCTGTTTA 1500
 DB 1441 ATTATAGTGGCAATTTTAGTAAATAGAGCAGGATTCGTCCTGTTGCTCGTCTGTTTA 1500

QY 1501 CGTTCCTTAA 1509
 DB 1501 CGTTCCTTAA 1509

RESULT 2

AX476887 AX476887 1509 bp DNA linear PAT 12-AUG-2002
LOCUS Sequence 15 from Patent WO0212294.
DEFINITION AX476887
ACCESSION AX476887
VERSION AX476887.1 GI:22216151
KEYWORDS
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE

1. Aderson, E. and Bohnsack, J.

A group b streptococcus polypeptides nucleic acids and therapeutic compositions and vaccines thereof

Patent: WO 0212294-A 15 14-FEB-2002;

ST. JUDE CHILDREN'S RESEARCH HOSPITAL (US) ; University of Utah

Research Foundation (US)

FEATURES

source

1. 1509

/organism="Streptococcus agalactiae"

/mol_type="unassigned DNA"

/db_xref="taxon:1311"

1. 1509

/notes="unamed protein product"

/codon_start=1

/transl_table=11

/protein_id="CAD43673.1"

/db_xref="GI:22216152"

/db_xref="RENTREMBL:CAD43673"

/translation="MKKKMQLSLVSLAFGMVSPVPIAFNAETGTTIVQDTQKA
TKKDYKFAEDIDNVAIVSNDKGSYLIPQKEBYKASTDFNSLFTTTNGGRYV
TKKDTASANEIATWAKSISANTTPSVTTESSNDGTEVINVSQYGYYSSTVNGAV
IMVTSVPNATIEKNTDAWGGGKTDVQKTSVGDVVKYITVKNVAVNHGTEKV
YQVVIKDTWPSASVDLNEGSYEVTITDGSNITLTQSEKATGKYNLENNFTI
TIPWANTPTQNGANDDPFKGINITITTYTGVLSGAKPGSADLPENTNIATI
NPTSNDPQKQVTVDDQITIKKIDGSTKASLQGAIFVLKNATGGQFLNFNDNNVEM
GTEANATEYTGADGIIITGLKEGYTIVLVEKAPILGNLLDNSQKVLGDGADFTN
SDLLNVPTVNNKGTSLPSTGGTIGTTFIYIIGAILVIGIVLVARRLRS"

ORIGIN

Query Match 100.0%; Score 1509; DB 6; Length 1509;
Best Local Similarity 100.0%; Pred. No. 3e-238;
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAAAAGAAAATGATTCAATCGCTGTAGTGGCGAGTTTAGCATTTGGTATGGCTGTA 60
DB 1 ATGAAAAGAAAATGATTCAATCGCTGTAGTGGCGAGTTTAGCATTTGGTATGGCTGTA 60
QY 61 TCACCAAGTTACGCCGATAGCTTTTGGCGCTGAGACAGGGAACAATACAGTTCAAGTACT 120
DB 61 TCACCAAGTTACGCCGATAGCTTTTGGCGCTGAGACAGGGAACAATACAGTTCAAGTACT 120
QY 121 CAAAAGGCCCAACCTATAAGCATATAAAGTTTTCATCGAGAAATAGATAATGCAAT 180
DB 121 CAAAAGGCCCAACCTATAAGCATATAAAGTTTTCATCGAGAAATAGATAATGCAAT 180
QY 181 GTATCTGATTTCGAATAAAGATGAGCTTCTTATTTAATTCCTCAAGTAAAGAGCTGAG 240
DB 181 GTATCTGATTTCGAATAAAGATGAGCTTCTTATTTAATTCCTCAAGTAAAGAGCTGAG 240
QY 241 TATAAAGCTTCAACTGATTTTAAATTCCTTTTTCACGAACTACTAATGAGGAGGAACA 300
DB 241 TATAAAGCTTCAACTGATTTTAAATTCCTTTTTCACGAACTACTAATGAGGAGGAACA 300
QY 301 TATGTAACATAAAGATGAGCTGAGCAAAATGAGATTTCGACATGGGCTAAATCTATA 360
DB 301 TATGTAACATAAAGATGAGCTGAGCAAAATGAGATTTCGACATGGGCTAAATCTATA 360
QY 361 TCAGCTAATACTACACAGTTTCCACTGTTTACTGAGTCAAAATAGATGTTACTGAGTT 420
DB 361 TCAGCTAATACTACACAGTTTCCACTGTTTACTGAGTCAAAATAGATGTTACTGAGTT 420

QY 421 ATTAATGTTTCCCAATATCGATATATTATGTTTCTAGCACTGTTATTAATGAGCTGTA 480
DB 421 ATTAATGTTTCCCAATATCGATATATTATGTTTCTAGCACTGTTATTAATGAGCTGTA 480
QY 481 ATTATGTTTACATCTGTAATCTCAAAATGCTACTATTATCATGAAAAGAAATCTGATGCGACA 540
DB 481 ATTATGTTTACATCTGTAATCTCAAAATGCTACTATTATCATGAAAAGAAATCTGATGCGACA 540
QY 541 TGGGAGATGTTGGTGGAAAAAATCTAGATCAAAAAACGTAATCTCGGTTGGTGTATACAGTC 600
DB 541 TGGGAGATGTTGGTGGAAAAAATCTAGATCAAAAAACGTAATCTCGGTTGGTGTATACAGTC 600
QY 601 AAATATATCTATTACTTATTAAGATGCGACATCAATATCATGTTTACAGAAAAGTGTATCAA 660
DB 601 AAATATATCTATTACTTATTAAGATGCGACATCAATATCATGTTTACAGAAAAGTGTATCAA 660
QY 661 TATGTTTATAAAGGATACTATGTCATCTGTTCTGTAGTTGATTTGAACGAAGGCTCTTAT 720
DB 661 TATGTTTATAAAGGATACTATGTCATCTGTTCTGTAGTTGATTTGAACGAAGGCTCTTAT 720
QY 721 GAAGTAATCTATTACTGATGATCAGGGAATATTAACAATCTTAATCAAGTTTCGGAATAA 780
DB 721 GAAGTAATCTATTACTGATGATCAGGGAATATTAACAATCTTAATCAAGTTTCGGAATAA 780
QY 781 GCAACTGGGAGATTAACCTGTTTAGAGGAAATAAATAATTTTCAAGTACTATTTCGCTGG 840
DB 781 GCAACTGGGAGATTAACCTGTTTAGAGGAAATAAATAATTTTCAAGTACTATTTCGCTGG 840
QY 841 GCAGCTTACCAATACTCTCAACCGGAAATATCTCAAAATGGAGCTAATGATGACTTTTTTAT 900
DB 841 GCAGCTTACCAATACTCTCAACCGGAAATATCTCAAAATGGAGCTAATGATGACTTTTTTAT 900
QY 901 AAGGGAATAAATAACAATCAAGTCACTTATACAGGAGTATTAAGAGTGGAGCTTAACCA 960
DB 901 AAGGGAATAAATAACAATCAAGTCACTTATACAGGAGTATTAAGAGTGGAGCTTAACCA 960
QY 961 GGTTCAGCTGATTTACAGAAAATAACAATTCGCAACCATCAACCCCATCTAGCAAT 1020
DB 961 GGTTCAGCTGATTTACAGAAAATAACAATTCGCAACCATCAACCCCATCTAGCAAT 1020
QY 1021 GATGACCCAGGTCAAAAAGTAAACAGTGGAGGATGGTCAAAATTAATAAAAAAATGAT 1080
DB 1021 GATGACCCAGGTCAAAAAGTAAACAGTGGAGGATGGTCAAAATTAATAAAAAAATGAT 1080
QY 1081 GGTTCACAAAAGCTTCATTATCAAGGTGCTATATTTGTTTAAAGAAATGCTACGGGTCAA 1140
DB 1081 GGTTCACAAAAGCTTCATTATCAAGGTGCTATATTTGTTTAAAGAAATGCTACGGGTCAA 1140
QY 1141 TTTCTAAACCTTTAAACGATACAAATTAACGTTGAATGGGGCACAAGCTAATGCAACAGAA 1200
DB 1141 TTTCTAAACCTTTAAACGATACAAATTAACGTTGAATGGGGCACAAGCTAATGCAACAGAA 1200
QY 1201 TATACAAACAGGACGATGCTATAATTCACATTACAGGCTTGAAGAGGAGTACATCTAT 1260
DB 1201 TATACAAACAGGACGATGCTATAATTCACATTACAGGCTTGAAGAGGAGTACATCTAT 1260
QY 1261 CTAGTTGAGAAAAGGCTCCCTTTAGGTTTACAAATTTGTTAGATAAATCTCAGAGGTTATT 1320
DB 1261 CTAGTTGAGAAAAGGCTCCCTTTAGGTTTACAAATTTGTTAGATAAATCTCAGAGGTTATT 1320
QY 1321 TTAGGAGATGGAGCCACTGATACGACTAATTCAGATAACCTTTTAGTTAACCCACTGTT 1380
DB 1321 TTAGGAGATGGAGCCACTGATACGACTAATTCAGATAACCTTTTAGTTAACCCACTGTT 1380
QY 1381 GAAAAATAACAAAGGTACTGAGTTGCCCTTCAACAGGTGGTATTGGTACAAATTTTCTAC 1440
DB 1381 GAAAAATAACAAAGGTACTGAGTTGCCCTTCAACAGGTGGTATTGGTACAAATTTTCTAC 1440
QY 1441 ATTATAGGTGCAATTTTATGTAATAGGAGCAAGGTATCGTGTGTTGCTCGTCGCTGTTA 1500
DB 1441 ATTATAGGTGCAATTTTATGTAATAGGAGCAAGGTATCGTGTGTTGCTCGTCGCTGTTA 1500

460 ACTG

ATAATGGAGCTGTAAATTATGGTTA

Figure 1

RESULT 10	AC007926	155204 bp	linear	HTG 17-JUL-2001
AC007926/c	Trypanosoma brucei chromosome II clone RPC193-3H15,			
LOCUS	IN PROGRESS ***			
DEFINITION				
ACCESSION	AC007926			
VERSION	AC007926.8	GI:114787210		
KEYWORDS	HTG; HTGS PHASE2.			

QY 650 AAGTGATCAATATGTTATTAAGGATACATGCGCTCTCTGTAGTTGATTTGAACG 709
Db 80837 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
QY 710 AAGGCTCTTATGAAGTAACATTTACTGATGGATCAGGGAATATTTCAACACTCAACTCAAG 769
Db 80777 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
QY 770 GTTCGGAAGAACACTGGGAAGTATACCTGTTAGAGGAAATAATAATTTTCACGATTA 829
Db 80717 ATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
QY 830 CTATTCGCTGGCGAGCTACCAATCTCCACCGGAAATACCTCAAAATGGAGCTAAATGATG 889
Db 80657 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
QY 890 ACTTTTTTTTATAA 902
Db 80597 ATGATAATAATAA 80585

RESULT 11

AL929250/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 94534)

Almeida.J.

Direct Submission

Submitted (16-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Sep 16, 2003 this sequence version replaced gi:27848021.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zf1sh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep

from a CHORI-211 BAC library

VECTOR: pTRABAC2.1

Clone-derived Zebrafish pUC subclones occasionally display

inconsistency over the length of mononucleotide A/T runs and

conserved TA repeats. Where this is found the longest good quality

representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat

discovery system (Zhihong Bao and Sean Eddy, submitted), and those

FEATURES

source

1. .94534

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clone="CH211-137A2"

/clone_lib="CHORI-211"

ORIGIN

Query Match 5.5%; Score 82.4; DB 5; Length 94534;

Best Local Similarity 44.5%; Pred.No. 0.00013;

Matches 326; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

QY 103 ATTACAGTTTCAAGATACCTCAAAAAGGCGCAACCTATAAAGCATATAAAGTTTTTGATGCA 162

Db 30450 AATATAGTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 163 GAAATAGATAATGCAAAATGATCTGATTCGATTCGATAAAGATGGAGCTTCTTTATTTATTCCT 222

Db 30390 AATAATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT

QY 223 CAAAGGTAAAGAACTCAGTATTAAGCTTCAACTGATTTTAAATTTCTCTTTTACGACAACT 282

Db 30330 AATAATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT

QY 30330 AATAATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT

Db 283 ACTAATGGAGGAGAAACATATGTAACCTAAAAAGATACCTCGTCAGCAAAATGAGATGCG 342

Db 30270 AATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 343 ACATGGGTAAATCTATATCAGCTAATACTACACCACTTTTCCACTGTTTCTGATGATGATAAT

Db 30210 AATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 403 AATGATGGTACTCAGGTATTATTATGTTTCCCAATATGATATTATTATTATTATTATTATTATT

Db 30150 AATAATGGTGATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA

QY 463 GTTAAATAATGGAGCTGTAATTTATGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAA 522

Db 30090 AGTAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 523 AAGAATACTGATCGGACATGCGGAGATGGTGGGAAAACTGTAGATCAAAAAACCTAC 582

Db 30030 AATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 583 TCGGTTGGTGATACAGTCAATATATCTATTACTTATAAGATGCAGTCAATTTATCATGGT 642

Db 29970 AATAATAATGGTGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 643 ACAGAAAAAGTGATCAATATGTTTATAAAGGATACCTATGCTCTCTCTGTAGTTGAT 702

Db 29910 AATA

QY 703 TTGAACCAAGGGTCTTATGAAGTAACTATTACTTATGATGATGATGATGATGATGATGATGAT

Db 29850 AGTAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 763 ACTCAAGGTCGGAAGAAAGCACTGGGAGTATAAAGCTGTTAGAGGAAATAATAATAATTC 822

Db 29790 AATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 823 ACCATTACTATT 834

Db 29730 AATAATTATTAA 29719

RESULT 12

PFMAL13.07

WPCOMMENT

Sequence split into 28 fragments LOCUS PFMAL13 Accession AL844509

Db 43595 ATAAATCATATAAATAATTATTTATATAATAACAATAAATGAGAAAAACCGAAAAAG 43655

QY 737 ATGGATCAGGAATATTCAACTTAACTCAAGGTTCGAAAACCACTGGGAAGTATA 796

Db 43655 AAGAAGTATCATTAATATATATCTGAAACCCCAATGCTATAGGAACATAATATGGTAATT 43714

QY 797 ACCTGTTAGAGGAAAATAATAATTCCA 823

Db 43715 ATATTTCCCAGGATGAAGATAATTACA 43741

RESULT 13	
BX537105	225581 bp DNA linear HTG 06-AUG-2003
LOCUS	Danio rerio clone DREY-3L24, 18 unordered pieces.
DEFINITION	
ACCESSION	BX537105
VERSION	BX537105.2 GI:31559379
KEYWORDS	HTG; HIGS PHASE1; HIGS CANCELLED.
SOURCE	Danio rerio (zebrafish)
ORGANISM	Danio rerio
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
	Cypriniformes; Cyprinidae; Danio.
REFERENCE	1 (bases 1 to 225581)
AUTHORS	Burton,J.
TITLE	Direct Submission
JOURNAL	Submitted (05-AUG-2003) Wellcome Trust Sanger Institute, Hinxtion, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfih-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jun 9, 2003 this sequence version replaced gi:31441851.
COMMENT	----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: zfih-help@sanger.ac.uk ----- Project Information Center project name: zk3L24 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Chemistry: Dye-terminator; 100% of reads Consensus quality: 217427 bases at least Q40 Consensus quality: 219130 bases at least Q30 Consensus quality: 220818 bases at least Q20 Insert size: 223881; sum-of-contigs Insert size: 219000; 7.3% error; agarose-fp Quality coverage: 6.82x in Q20 bases; sum-of-contigs Quality coverage: 7.43x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 18 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 20502: contig of 20602 bp in length * 20603 20702: gap of 100 bp * 20703 57831: contig of 37129 bp in length * 57832 57931: gap of 100 bp * 57932 65164: contig of 7233 bp in length * 65165 65264: gap of 100 bp * 65265 71392: contig of 6128 bp in length * 71393 71492: gap of 100 bp * 71493 93677: contig of 22185 bp in length * 93678 93777: gap of 100 bp * 93778 101598: contig of 7821 bp in length * 101599 101699: gap of 100 bp * 101699 128689: contig of 26991 bp in length * 128690 128789: gap of 100 bp * 128790 138632: contig of 9843 bp in length * 138633 138732: gap of 100 bp * 138733 144193: contig of 5461 bp in length

```
* 144194 144293: gap of 100 bp
* 144294 155818: contig of 11525 bp in length
* 155819 15918: gap of 100 bp
* 15919 168354: contig of 12436 bp in length
* 168355 168454: gap of 100 bp
* 168455 171301: contig of 2847 bp in length
* 171302 171401: gap of 100 bp
* 171402 180364: contig of 8963 bp in length
* 180365 180464: gap of 100 bp
* 180465 188604: contig of 8140 bp in length
* 188605 188704: gap of 100 bp
* 188705 192855: contig of 4151 bp in length
* 192856 192955: gap of 100 bp
* 192956 201755: contig of 8799 bp in length
* 201756 201854: gap of 100 bp
* 201855 211372: contig of 9518 bp in length
* 211373 211472: gap of 100 bp
* 211473 225581: contig of 14109 bp in length.
FEATURES
    source
        1..225581
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7965"
            /clone="DKEY-3L24"
            /clone_lib="DanioKey"
            1..20602
                /notes="assembly fragment:01463"
                fragment_chain:1
                20703..57831
                    /notes="assembly fragment:01892"
                    fragment_chain:1
                    57932..65164
                        /notes="assembly fragment:02319"
                        fragment_chain:1
                        65265..71392
                            /notes="assembly fragment:02166"
                            fragment_chain:1
                            71493..93677
                                /notes="assembly fragment:01041"
                                fragment_chain:1
                                93778..101598
                                    /notes="assembly fragment:02235"
                                    fragment_chain:1
                                    101699..128689
                                        /notes="assembly fragment:01236"
                                        fragment_chain:1
                                        128790..138632
                                            /notes="assembly fragment:02758"
                                            fragment_chain:1
                                            138733..144193
                                                /notes="assembly fragment:02245"
                                                fragment_chain:1
                                                144294..155918
                                                    /notes="assembly fragment:01852"
                                                    fragment_chain:1
                                                    155919..168354
                                                        /notes="assembly fragment:01722"
                                                        fragment_chain:1
                                                        168455..171301
                                                            /notes="assembly fragment:00772"
                                                            fragment_chain:2
                                                            171402..180364
                                                                /notes="assembly fragment:02453"
                                                                fragment_chain:2
                                                                180465..188604
                                                                    /notes="assembly fragment:00524"
                                                                    fragment_chain:2
                                                                    188705..192855
                                                                        /notes="assembly fragment:02841"
                                                                        fragment_chain:2
                                                                        192956..201754
                                                                            /notes="assembly fragment:00644"
                                                                            fragment_chain:2
```

```
misc_feature 201855..211372
              /notes="assembly fragment:02536"
              fragment_chain:2
misc_feature 211473..225581
              /notes="assembly fragment:01373"
              fragment_chain:2
ORIGIN
Query Match      5.3%; Score 80.4; DB 2; Length 225581;
Best Local Similarity 45.3%; Pred. No. 0.00021;
Matches 291; Conservative 0; Mismatches 351; Indels 0; Gaps 0;
QY 261 TAATTCTCTTTTACGCAACTACTTAATCGAGGAGACATATGTACTACTAAAAAGATAC 320
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151722 TACTTCTACTACTACTACTACTTATATAATAATAATAATAATAATAATAATAATGA 151781
QY 321 TCGCTCAGCAAAATGAGATTGCGACATCGGCTAAATCTATATCAGTAAATACACACAGT 380
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151782 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGA 151841
QY 381 TTCCACTGTTACTGAGTCAAAATAATGATGGTACTGAGGTTATTAAATGTTCCCAATATGG 440
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151842 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 151901
QY 441 ATATTATTATGTTCTTAGCACCTGTTAATATGAGGCTGTAATATGCTTACATCTGTAAC 500
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151902 TAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 151961
QY 501 TCCAAATGCTACTATTTCATGAAAGAAATACTGATGCGACATGGGAGATGGTGTGAAA 560
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151962 TATTGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 152021
QY 561 AACTGTAGATCAAAAACGCTACTCGGTTGGTGATACAGTCAAAATATATCTATTCTTATAA 620
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152022 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 152081
QY 621 GATGTCAGTCAATTATCATGGTACAGAAAAGTGATCATATGTTATAAGGATACCTAT 680
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152082 TAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 152141
QY 681 GCCATCTGCTTCTGTAGTTGATTGAACGAGGCTTTATGAAGTAACTACTATTCTGATGG 740
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152142 TAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 152201
QY 741 ATCAGGAATATTACAACTCTAATCAAGTTCGGAAGAAAGCACTGGGAAGTATAACCT 800
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152202 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 152261
QY 801 GTTAGAGGAAATAATAATTTCCAGATTACTATTCGCTGGGAGCTACCAATCTCCAAC 860
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152262 TAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 152321
QY 861 CGGAAATACCTCAAAATGGAGCTAATGATGACTTTTTTTTATAA 902
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152322 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 152363
```

```
RESULT 14
AF397467/c 10115 bp DNA linear VRT 27-NOV-2001
LOCUS Ictalurus punctatus novel immune-type receptor 2 (NITR2) and novel
DEFINITION Immune-type receptor 4 (NITR4) genes, complete cds.
ACCESSION AF397467
VERSION AF397467.1 GI:17105096
KEYWORDS Ictalurus punctatus (channel catfish)
SOURCE Ictalurus punctatus
ORGANISM Ictalurus punctatus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
          Ictaluridae; Ictalurus.
REFERENCE 1 (bases 1 to 10115)
AUTHORS Hawke,N.A., Yoder,J.A., Haire,R.N., Mueller,M.G., Litman,R.T.,
          Miracile,A.L., Stuge,T., Shen,L., Miller,N. and Litman,G.W.
```


QY	965	CAGCTGATTACCAGAAATACAAACATTGGACCATCAACCCCAATACTAGCAATGATG	1024
Db	166410	ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG	166351
QY	1025	ACCCAGGTCAAAAAGTAAACAGTGAGGGATGGTCAAATTTACTATATAAAAAAATTTGATGGT	1083
Db	166350	ATACAGATACAGATACAGGATTTGGAGCTGGAGTCATTTAAAGGAGTCAAGAGTGCAGGT	166292

Search completed: September 13, 2004, 08:14:20
Job time : 5991.67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 10:28:56 ; Search time 45 Seconds
(without alignments)
3493.409 Million cell updates/sec

Title: US-10-009-254-2

Perfect score: 2578

Sequence: 1 MKKMIQSLVSLAFGMVAV.....GALLVIGIVLVARRLRS 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/FCR_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2578	100.0	502	12	US-10-333-002-16
2	334	13.0	627	12	US-10-206-576-218
3	334	13.0	627	16	US-10-661-809-13
4	318	12.3	625	16	US-10-661-809-19
5	274.5	10.6	560	12	US-10-206-576-220
6	265.5	10.3	665	10	US-09-769-787-127
7	232	9.0	793	12	US-10-282-122A-60689
8	222.5	8.6	588	9	US-09-864-761-36047
9	219	8.5	560	12	US-10-333-002-19
10	219	8.5	915	14	US-10-193-764-35
11	219	8.5	1222	14	US-10-193-764-37
12	219	8.5	1228	14	US-10-193-764-34
13	218	8.5	1465	15	US-10-369-493-13955
14	212	8.2	645	12	US-10-282-122A-73634
15	212	8.2	2551	12	US-10-282-122A-74083
					Sequence 16, Appl
					Sequence 218, Appl
					Sequence 13, Appl
					Sequence 19, Appl
					Sequence 220, Appl
					Sequence 127, Appl
					Sequence 60689, A
					Sequence 36047, A
					Sequence 19, Appl
					Sequence 35, Appl
					Sequence 19, Appl
					Sequence 37, Appl
					Sequence 34, Appl
					Sequence 13955, A
					Sequence 73634, A
					Sequence 74083, A

16	208.5	8.1	476	16	US-10-661-809-11	Sequence 11, Appl
17	200.5	7.8	886	10	US-09-769-787-126	Sequence 126, Appl
18	199	7.7	473	16	US-10-661-809-17	Sequence 17, Appl
19	198.5	7.7	724	12	US-10-282-122A-45795	Sequence 45795, A
20	198	7.7	596	10	US-09-946-374-243	Sequence 243, Appl
21	198	7.7	596	12	US-10-206-915-310	Sequence 310, Appl
22	198	7.7	596	12	US-10-199-670-310	Sequence 310, Appl
23	198	7.7	596	12	US-10-201-858-310	Sequence 310, Appl
24	198	7.7	596	12	US-10-205-890-310	Sequence 310, Appl
25	198	7.7	596	12	US-10-208-024-310	Sequence 310, Appl
26	198	7.7	596	12	US-10-201-853-310	Sequence 310, Appl
27	198	7.7	596	12	US-10-063-745-100	Sequence 100, Appl
28	198	7.7	596	12	US-10-063-512-100	Sequence 100, Appl
29	198	7.7	596	12	US-10-063-513-100	Sequence 100, Appl
30	198	7.7	596	12	US-10-063-515-100	Sequence 100, Appl
31	198	7.7	596	12	US-10-063-549-100	Sequence 100, Appl
32	198	7.7	596	12	US-10-063-569-100	Sequence 100, Appl
33	198	7.7	596	12	US-10-063-551-100	Sequence 100, Appl
34	198	7.7	596	12	US-10-174-581-310	Sequence 310, Appl
35	198	7.7	596	12	US-10-176-483-310	Sequence 310, Appl
36	198	7.7	596	12	US-10-176-749-310	Sequence 310, Appl
37	198	7.7	596	12	US-10-176-914-310	Sequence 310, Appl
38	198	7.7	596	12	US-10-176-915-310	Sequence 310, Appl
39	198	7.7	596	12	US-10-006-485A-243	Sequence 243, Appl
40	198	7.7	596	12	US-10-013-907A-243	Sequence 243, Appl
41	198	7.7	596	12	US-10-015-499A-243	Sequence 243, Appl
42	198	7.7	596	12	US-10-063-555-100	Sequence 100, Appl
43	198	7.7	596	12	US-10-063-563-100	Sequence 100, Appl
44	198	7.7	596	12	US-10-063-594-100	Sequence 100, Appl
45	198	7.7	596	12	US-10-063-553-100	Sequence 100, Appl
46	198	7.7	596	12	US-10-063-554-100	Sequence 100, Appl
47	198	7.7	596	12	US-10-176-484-310	Sequence 310, Appl
48	198	7.7	596	12	US-10-180-550-310	Sequence 310, Appl
49	198	7.7	596	12	US-10-183-014-310	Sequence 310, Appl
50	198	7.7	596	12	US-10-187-738-310	Sequence 310, Appl
51	198	7.7	596	12	US-10-187-740-310	Sequence 310, Appl
52	198	7.7	596	12	US-10-187-883-310	Sequence 310, Appl
53	198	7.7	596	12	US-10-194-363-310	Sequence 310, Appl
54	198	7.7	596	12	US-10-194-460-310	Sequence 310, Appl
55	198	7.7	596	12	US-10-194-463-310	Sequence 310, Appl
56	198	7.7	596	12	US-10-194-484-310	Sequence 310, Appl
57	198	7.7	596	12	US-10-195-884-310	Sequence 310, Appl
58	198	7.7	596	12	US-10-195-896-310	Sequence 310, Appl
59	198	7.7	596	12	US-10-196-744-310	Sequence 310, Appl
60	198	7.7	596	12	US-10-196-755-310	Sequence 310, Appl
61	198	7.7	596	12	US-10-196-757-310	Sequence 310, Appl
62	198	7.7	596	12	US-10-197-704-310	Sequence 310, Appl
63	198	7.7	596	12	US-10-197-710-310	Sequence 310, Appl
64	198	7.7	596	12	US-10-198-758-310	Sequence 310, Appl
65	198	7.7	596	12	US-10-198-766-310	Sequence 310, Appl
66	198	7.7	596	12	US-10-199-304-310	Sequence 310, Appl
67	198	7.7	596	12	US-10-199-309-310	Sequence 310, Appl
68	198	7.7	596	12	US-10-199-313-310	Sequence 310, Appl
69	198	7.7	596	12	US-10-199-456-310	Sequence 310, Appl
70	198	7.7	596	12	US-10-201-329-310	Sequence 310, Appl
71	198	7.7	596	12	US-10-202-412-310	Sequence 310, Appl
72	198	7.7	596	12	US-10-206-919-310	Sequence 310, Appl
73	198	7.7	596	12	US-10-206-922-310	Sequence 310, Appl
74	198	7.7	596	12	US-10-206-924-310	Sequence 310, Appl
75	198	7.7	596	12	US-10-206-928-310	Sequence 310, Appl
76	198	7.7	596	12	US-10-207-314-310	Sequence 310, Appl
77	198	7.7	596	12	US-10-207-921-310	Sequence 310, Appl
78	198	7.7	596	12	US-10-207-922-310	Sequence 310, Appl
79	198	7.7	596	12	US-10-208-027-310	Sequence 310, Appl
80	198	7.7	596	12	US-10-013-910A-243	Sequence 243, Appl
81	198	7.7	596	12	US-10-174-570-310	Sequence 310, Appl
82	198	7.7	596	12	US-10-183-005-310	Sequence 310, Appl
83	198	7.7	596	12	US-10-236-254A-243	Sequence 243, Appl
84	198	7.7	596	12	US-10-015-395A-243	Sequence 243, Appl
85	198	7.7	596	13	US-10-006-867-100	Sequence 100, Appl
86	198	7.7	596	13	US-10-052-586-310	Sequence 310, Appl
87	198	7.7	596	13	US-10-063-547-100	Sequence 100, Appl
88	198	7.7	596	14	US-10-174-590-310	Sequence 310, Appl

89 198 7.7 596 14 US-10-176-758-310 Sequence 310, App
90 198 7.7 596 14 US-10-175-737-310 Sequence 310, App

ALIGNMENTS

RESULT 1
US-10-333-002-16
; Sequence 16, Application US/10333002
; Publication No. US20040071729A1
; GENERAL INFORMATION:
; APPLICANT: Adderson, Elisabeth
; TITLE OF INVENTION: Group B Streptococcus Polypeptides Nucleic Acids and Therapeutic
; FILE REFERENCE: 2511-1-001 (SU-0039)
; CURRENT APPLICATION NUMBER: US/10/333,002
; PRIOR FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: PCT/US01/24795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/634,341
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-10-333-002-16

Query Match 100.0%; Score 2578; DB 12; Length 502;
Best Local Similarity 100.0%; Pred. No. 1.3e-198;
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKKKMIQSLVLASLAFGMVSPVTPPIAFAAETGTTIVQDTQKGATYKAYKVFDAEIDNAN 60
DB 1 MKKKMIQSLVLASLAFGMVSPVTPPIAFAAETGTTIVQDTQKGATYKAYKVFDAEIDNAN 60
QY 61 VSDSNKDGASLYLPQGEAEYKASTDFNSLPTTTTNGGRTYVTKDTSANETATWAKSI 120
DB 61 VSDSNKDGASLYLPQGEAEYKASTDFNSLPTTTTNGGRTYVTKDTSANETATWAKSI 120
QY 121 SANTTPVSTVTSNNDGTEVINVSQYGYVSVTVNNGAVIMVTSVTPNATHEKNTDAT 180
DB 121 SANTTPVSTVTSNNDGTEVINVSQYGYVSVTVNNGAVIMVTSVTPNATHEKNTDAT 180
QY 181 WDGSGGKTVDQKTSYSGDTPVKYITTYKNAVNHGTEKTYQYVTKDTPMPASVVDLNEGSY 240
DB 181 WDGSGGKTVDQKTSYSGDTPVKYITTYKNAVNHGTEKTYQYVTKDTPMPASVVDLNEGSY 240
QY 241 EVTITDGSNITTLTGSEKATGKYNLENNNFTTIPWAAATNTPTGNTONGANDDFY 300
DB 241 EVTITDGSNITTLTGSEKATGKYNLENNNFTTIPWAAATNTPTGNTONGANDDFY 300
QY 301 KGINTITVTYTVGLKSGKPGSADLPENTNIATINPNTSNDPQKVTVRVDGQITIKKID 360
DB 301 KGINTITVTYTVGLKSGKPGSADLPENTNIATINPNTSNDPQKVTVRVDGQITIKKID 360
QY 361 GSTKASLQGAIFVLKNATGFLNPDNTNVEWGTTEANATYTTGADGIIITLKGTTY 420
DB 361 GSTKASLQGAIFVLKNATGFLNPDNTNVEWGTTEANATYTTGADGIIITLKGTTY 420
QY 421 LVKKAPLGNLNDNSOKVILGQATDTTNSDNLNVPNTVENNKGTELSTGGIGTIFY 480
DB 421 LVKKAPLGNLNDNSOKVILGQATDTTNSDNLNVPNTVENNKGTELSTGGIGTIFY 480
QY 481 IIGAILVIGAGIVLVARRRLLRS 502
DB 481 IIGAILVIGAGIVLVARRRLLRS 502

RESULT 2

US-10-206-576-218
; Sequence 218, Application US/10206576
; Publication No. US20030017495A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/206,576
; FILING DATE: 29-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/071,035
; FILING DATE: 1998-05-04
; APPLICATION NUMBER: US 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: US 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: US 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB369P1D1
; INFORMATION FOR SEQ ID NO: 218:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 627 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 218:
US-10-206-576-218

Query Match 13.0%; Score 334; DB 12; Length 627;
Best Local Similarity 26.3%; Pred. No. 5e-18;
Matches 150; Conservative .79; Mismatches 223; Indels 118; Gaps 26;
QY 3 KKMIOQLVLASLAFGMVSPVTPPIAFAA--ETGTTIVQDTQKGATYKAYKVFDAEIDNAN 60
DB 100 KQAVQSL-----TPGKPVAGITTDANGVTVQLPKKQNGKDAVYTIKEPKGV 148
QY 61 VSDSNKDGAA--SYLIPQGEAEYKASTDFNSL-----TTTTNGGRTYVTKDTSANP- 112
DB 149 VAATNNWAFPPVYEMIKOTDGSYKYGTEBELAVHYHPKNNVANDGSLHVKYGTAE--NEG 207
QY 113 -----TATWA-----KSIANTTPV--STVTESNND 136
DB 208 LMGAEFVTSKSGSPGVKVIQGVKGLTWTTKDQAKRFITGKSYEIGENDPTEAEN- 266
QY 137 GTEVINVS--QYGYTVVS--STVNNGAVIMVTSVTP-----NATHEK--NTDATWGDG 184
DB 267 GTGELTVKNLEVGSYLIEVKAPNNAELIENQTKTFTIEANNQTPVEKTVKNDTSKVDK 326
QY 185 GSKTVDDQKTSYSGDTPVKYITTYKNAVNHGTE-----KVQYVTKDTPMPASVVD--LNEG 239
DB 327 TTPSLDGDKDAIGEKIKYQISVNIPLGIADKGDANKYKFNLDKDAALTFDNTVTSGE 386
QY 240 YEVTITDGSNITTLTGSEKATGKYNLENNNFTTIT--PWAATNTPTGNTONGANDD 297
DB 387 YAYALYDGDVTI-----APENYQVTEQANGFTVAHPAYIPTLTPGG----- 428